

## Research Article

# Discovering Congested Routes Using Vehicle Trajectories in Road Networks

Kyoung Soo Bok,<sup>1</sup> He Li,<sup>2</sup> Jong Tae Lim,<sup>1</sup> and Jae Soo Yoo<sup>1</sup>

<sup>1</sup> School of Information and Communication Engineering, Chungbuk National University, 52 Naesudong-ro, Chungbuk, Seowon-Gu, Cheongju 362-763, Republic of Korea

<sup>2</sup> School of Software, Xidian University, No. 2 South Taibai Road, Xi'an City, Shaanxi 710071, China

Correspondence should be addressed to Jae Soo Yoo; [yjs@chungbuk.ac.kr](mailto:yjs@chungbuk.ac.kr)

Received 28 August 2014; Accepted 23 October 2014

Academic Editor: Seungmin Rho

Copyright © 2015 Kyoung Soo Bok et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The popular route recommendation and traffic monitoring over the road networks have become important in the location-based services. The schemes to find out the congested routes were proposed by considering the number of vehicles in a road segment. However, the existing schemes do not consider the features of each road segment such as width, length, and direction in a road network. Furthermore, the existing schemes fail to consider the average moving speed of vehicles. Therefore, they can detect the incorrect density routes. To overcome such problems, we propose a new discovering scheme of congested routes through the analysis of vehicle trajectories in a road network. The proposed scheme divides each road into segments with different width and length in a road network. And then, the congested road segment is detected through the saturation degree of the road segment and the average moving speed of vehicles in the road segment. Finally, we compute the final congested routes by using a clustering scheme. The experimental results have shown that the proposed scheme can efficiently discover the congested routes in the different directions of the roads.

## 1. Introduction

Recently, with the increase of the use of mobile devices, the location-based services are becoming increasingly popular. Since the rapidly increased satellites and GPS (global position system) technologies have developed, it is possible to collect a large amount of trajectory data of moving objects such as the vehicle position data, hurricane track data, and animal movement data [1–4]. The analysis over these trajectory data is becoming important for many applications, such as meteorological observation and forecast, animal habits observation, road traffic situation analysis, and navigation in transportations [5–9]. According to the recorded trajectory data and road networks, the moving pattern, traffic situation, and road recommendation services can be supported [1, 2, 10–12].

Recently, with the continuously increasing vehicles, the route recommendation service is becoming more and more important [1, 4, 7, 8, 13–15]. For road network based applications, the mobility of the vehicle is road network constrained.

Most of the existing schemes try to monitor and forecast the traffic by using the recorded history trajectory data of vehicles equipped with GPS devices. The index based schemes construct an index by adopting the trajectory data of the vehicles [5, 6]. And then the routes are recommended according to the history trajectory data of the related vehicles. The clustering based schemes generate the density routes of the road networks by analyzing the trajectory data of vehicles [9, 13, 16, 17]. Reference [1] proposed the MPR scheme for discovering the popular route between two locations by observing the traveling behaviors of many previous users. Reference [18] studied the problem of finding reasonable alternative routes in road networks. In NETSCAN scheme [13], the density routes of the road networks are evaluated by considering both the location and time of the vehicles.

According to the trajectory data, the number of the vehicles within a specific road segment and time period can be used to identify the density routes of a road. However, the existing schemes are not suitable for real road networks. There are three major problems of the existing schemes:

(1) the directions of the roads in the road networks are not considered; (2) the widths and lengths of the road segments are not considered; and (3) the average moving speed of vehicles within a road segment is not considered. In the real road network environments, each road is divided into two directions: positive direction and negative direction. The vehicles in the road toward different directions do not affect each other. Therefore, the congested routes are inaccurate if the directions of the road are not considered. The width and length of each road segment are different in a road network, which will also affect the accuracy of the congested routes. Furthermore, the average moving speed of vehicles within a road segment can identify the congestion of the road.

To overcome these problems, we propose a congested routes discovering scheme, in which the directions, width, and length of roads are considered for congested route discovering in real road network environments. The proposed scheme divides the roads into segments with different widths and lengths. And then, the congested road segments are extracted by considering the average moving speed of the vehicles and the saturation degree of each road segment in the road networks. Finally, we perform clustering schemes to compute the final congested routes in the whole road networks.

The remainder of the paper is organized as follows. We will discuss related work in Section 2. Section 3 presents the details of the proposed scheme. Section 4 contains experimental evaluation that demonstrates the superiority of our proposed scheme. Finally, Section 5 concludes this paper.

## 2. Related Work

Discovering hot routes or popular routes is a very important technique to support location-based service in road network. Reference [1] studied the discovering scheme of the most popular route (MPR) between two locations by observing the traveling behaviors of many previous users. Reference [1] presented an algorithm to establish the transfer network model for a collection of historical trajectories and utilized the absorbing Markov chain model to derive the transfer probability for transfer nodes. Maximum probability product algorithm is used for discovering the MPR from a transfer network based on the popularity indicators in a breadth-first scheme. Reference [16] defined a hot route which is a general path in the road network containing heavy traffic and studied the discovering scheme of hot routes. Reference [16] proposed a new density-based algorithm called FlowScan. It is a robust algorithm that can handle the complexities in the data and was verified through extensive experiments. Instead of clustering the moving objects, road segments are clustered based on the density of common traffic they share. Reference [15] studied the problem of vehicular traffic density estimation, utilizing the information cues present in the cumulative acoustic signal acquired from a roadside-installed single microphone.

Vehicles which are located in a congested area try to move to a noncongested area. Reference [10] proposed a route discovery method for alleviating traffic congestions to

provide a driving route whose trip time becomes short. The proposed method does not need global traffic information but regional traffic information for each vehicle. The vehicle calculates a route for a destination where a summation of evaluation values for roadway segments in the route becomes minimal. Given a spatial range and a user preference of depth/breadth specified by a user, [11] processed a pattern-aware trajectory search (PATS) to retrieve the top  $K$  trajectories passing through popular ROIs. PATS support trip planning without requiring prior knowledge of ROIs in the specified spatial range. PATS used a user movement graph to capture travel patterns hidden in trajectories and develop an algorithm to determine the attractive scores of the ROIs and proposed an algorithm BTS for efficiently retrieving the top  $K$  trajectories.

Reference [19] proposed a fast path algorithm of finding the best shortest paths in the road network to solve the path planning problem in route guidance systems in terms of accuracy and speed. Reference [20] proposed an adaptive fastest path algorithm capable of efficiently accounting for important driving and speed patterns mined from a large set of traffic data. Reference [20] presented a routing algorithm which uses the road hierarchy and precomputed areas to limit the search space. This improves trip duration by using upgraded roads whenever beneficial and finds routes that take into consideration both speed and driving patterns.

Reference [13] proposed a scheme for discovering dense routes by clustering similar road sections according to both traffic and location in each time period. Reference [13] presented NETSCAN which carries out the clustering of dense sections and incorporates them by forming dense routes. NETSCAN cluster the road sections based upon the network density statistics. This clustering takes into account the orientation of the trajectory. Besides, this method utilizes the network topology to create relevant clusters. To propose a model to assess the evolution for dense route pairs at two consecutive time intervals, DENSITYLINK algorithm is presented. DENSITYLINK allows the characterization of the evolution of the dense road network. Reference [17] proposed a time-based clustering algorithm called Tk-means that adapts the k-means algorithm for trajectory data. Tk-means cluster the objects based on the time intervals of different trajectory's motions. If an object spans different time intervals, it will eventually belong to different clusters. Tk-means used two approaches, an exact method and an approximate method. The exact method computes the actual clusters visited by the object throughout its life time and the approximate method exactly computes some of the actual visited clusters and, based on those computed clusters along with the clusters generated from the remaining data set, it predicts the future motion pattern of the query object.

## 3. The Proposed Scheme

*3.1. Data Model.* We assume that the road network is represented by a graph  $G(N, E)$ , where  $N$  denotes the node which is the intersection between different road segments and  $E$  denotes the edge which is used to connect two adjacent nodes

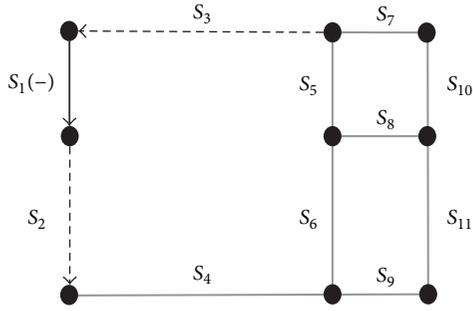


FIGURE 1: The data model of road network.

in the road network. The trajectory of a vehicle is represented by  $Tr$ . Each node  $N_i$  in the road network is represented by a point  $\{x_i, y_i\}$ .  $E_i$  denotes a segment of the road network. “+” and “-” are used to represent the different directions of vehicles in a road network. Moreover, since the length and the number of lane of each road are different, the length and width of each road segment are stored. Therefore, each road segment is represented by  $S_i(\pm) = \{N_i, N_j, \text{length}, \text{width}\}$ , where length is the length of a road segment and width is the number of traffic lanes. As shown in Figure 1,  $S_2$  and  $S_3$  are the neighbor segments of  $S_1$  in a road network  $G$ .

In a road network  $G$ , each road segment  $S_i$  stores the information of its directly connected road segments. This information is used for the following clustering evaluations. Since the vehicles may move continuously or stay in a position, it is necessary to have the location knowledge of each vehicle according to the timestamps. Suppose that the trajectory  $Tr$  of each vehicle is as follows:

$$Tr_n = \langle (S_1(\pm), T_1), (S_2(\pm), T_2), \dots, (S_k(\pm), T_k) \rangle, \quad (1)$$

where  $S_i$  denotes the segment ID and  $T_i$  is the timestamp. According to  $T_i$ , the location of each vehicle can be retrieved easily.

**3.2. Congested Road Segment.** The initial road information and trajectory data can be used to determine the existence of vehicles in each road segment of different directions. The complexity value of each road segment is evaluated. The road segment is considered as congested region if the evaluated complexity value is higher than the predefined threshold value. The clustering algorithm is performed among congested road segments. The road segments with low moving speed and high number of vehicles are determined as congested road segments. Figure 2 shows the procedure of computing congested routes.

In this paper, the congested road segments are computed according to the different directions of the roads. The location and direction of each vehicle can be retrieved from the recorded trajectory data. The complexity value of each road segment is computed by considering the average speed of the vehicles in the road segment and the saturation degree of the road segment. The fast moving speed indicates that the congestion of the road segment is low. In contrast, the low moving speed indicates that the congestion of the road

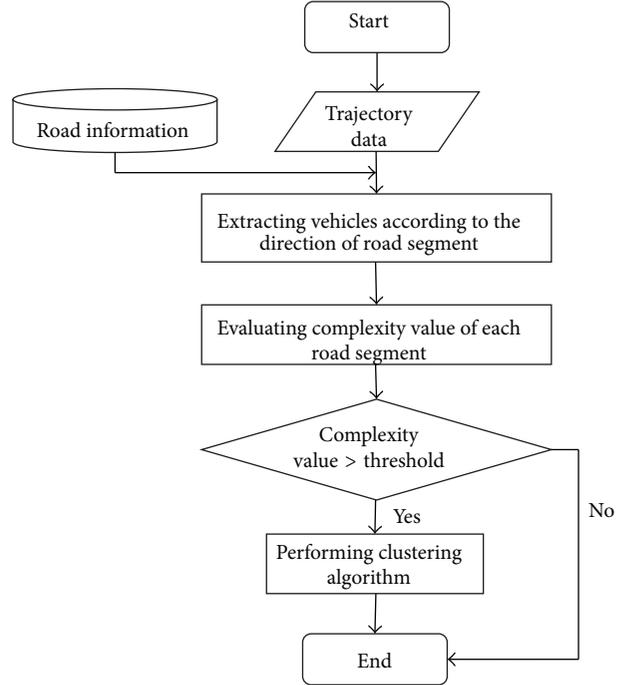


FIGURE 2: The procedure of computing congested routes.

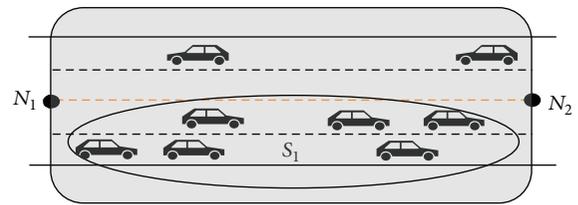


FIGURE 3: The road segment between nodes  $N_1$  and  $N_2$  in a road network.

segment is high. The saturation degree is computed based on the number of the vehicles within a road segment and the length and width of a road segment, which are indicated in Figure 3. We define that the congested road segments within a road network are the road segments with high complexity values.

The average moving speed ( $Av$ ) of the vehicles in a road segment according to different directions is computed by the following equation (2), where  $V(Ob_i)$  denotes the moving speed of a vehicle  $Ob_i$ . The saturation ( $Sat$ ) according to the width ( $S_{width}$ ) and length ( $S_{length}$ ) of a road segment is computed in (3), where  $Ob_n$  denotes the number of the vehicles in a road segment. As a result, the complexity value of a road segment is computed by (4) which combines (2) and (3);  $\alpha$  denotes the weight value between the average moving

```

Congested routes detection algorithm
Input:
Set of road segment list  $S = \{S_1, S_2, \dots, S_{n+1}\}$ 
Set of trajectory  $Tr = \{Tr_1, Tr_2, \dots, Tr_n\}$ 
Set of time interval  $T = \{T_1, T_2, \dots, T_n\}$ 
Threshold  $\beta$ —trajectory velocity and road weight
Output:
Threshold  $\alpha$ —road density
Set of cluster groups  $CH = \{CH_1, CH_2, \dots, CH_n\}$ 
 $CH \leftarrow$  Initialization
for every time interval  $t$ 
while there exist  $Tr$  in segment list  $S$ 
for each satisfied  $\alpha$  in  $S$ 
segment  $S =$  create candidate  $c$ 
end for
if exist neighbor segment  $N$  in candidate  $c$ 
create  $CH \leftarrow$  extend from  $c$  to  $N$ 
else if
create  $CH \leftarrow$  candidate  $c$ 
end if
end while
end for

```

ALGORITHM 1: The algorithm of the proposed scheme.

speed of vehicles and the saturation of a road segment. Consider

$$Av_i(\pm) = \frac{\sum_{i=1}^n V(Ob_i)}{Ob_n(\pm) \times V_{\max}}, \quad (2)$$

$$Sat_i(\pm) = 1 - \frac{Ob_n(\pm)}{S_{\text{width}} \times S_{\text{length}}}, \quad (3)$$

$$Ri_i(\pm) = \alpha \times Av_i(\pm) + (1 - \alpha) \times Sat_i(\pm). \quad (4)$$

**3.3. Congested Routes Detection.** In this section, we present the algorithm of congested routes detection. The algorithm operates in two phases. In the first phase, the complexity value of each road segment in the road network according to different directions is computed. In the second phase, the congested routes of a road network are evaluated by clustering the congested road segments with each time interval. When the complexity value of a road segment is larger than the predefined threshold value, the road segment is considered as a congested road segment. Finally, the congested road segments with same direction are clustered together. Algorithm 1 shows the congested routes detection algorithm.

Since the complexity values of road segments are changed according to different timestamps, they have to be computed periodically. The congested routes of a road network are computed according to the complexity values. Figure 4 shows the congested routes (the dotted areas) of a road network in different timestamp  $T$ . According to the recorded trajectory data of vehicles at time  $T = 0$ , the congested routes of the road network of different directions are generated, such as  $S_2(+)$  and  $S_5(-)$  in Figure 4(a). As shown in Figure 4(b), when time  $T = 1$ ,  $S_1(+)$ ,  $S_5(-)$ , and  $S_6(-)$  are evaluated as congested road

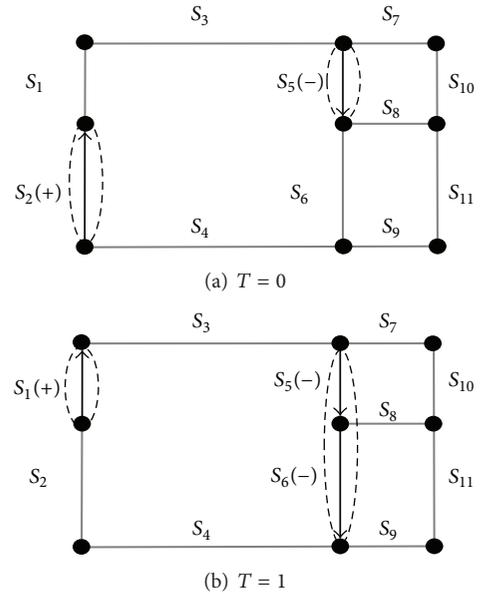


FIGURE 4: The clusters according to different timestamps.

segments. Since  $S_5(-)$  and  $S_6(-)$  are neighbor road segments and have the same direction in the road network, they are clustered together.

## 4. Performance Evaluation

In this section, we introduce the performance evaluation by comparing the proposed scheme with the existing scheme

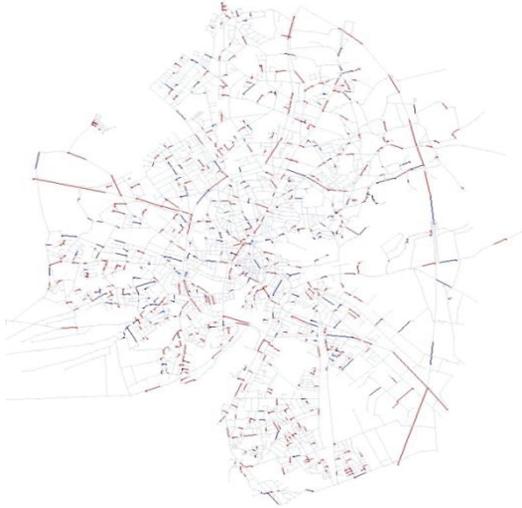


FIGURE 5: The congested route of Oldenburg city.

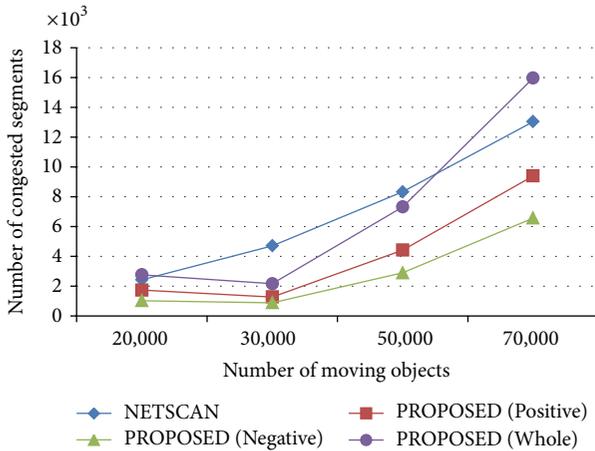


FIGURE 6: The number of congested routes according to the number of vehicles.

NETSCAN [13]. The vehicles are generated by the network-based generator [21]. The complexity values and clusters are generated according to the number of vehicles in each time interval. All of the experiments are coded in Java and the experiments are performed in Intel i3 3.0 GHz CPU and 4 G memory. Table 1 summarizes the parameters for this performance evaluation.

In the first experiment, we show the congested routes road networks of Oldenburg city by using our proposed scheme. In this experiment, the total number of vehicles in the road network is set to 50,000 and the saturation of the road is set to 30%. As shown in Figure 5, the results indicated that the congested routes of the road network are different according to the different directions of the road networks. The blue and red regions represent the congested routes of the road networks in positive direction and negative direction, respectively.

In Figure 6, we compare the NETSCAN scheme with our proposed scheme. The number of the congested routes of

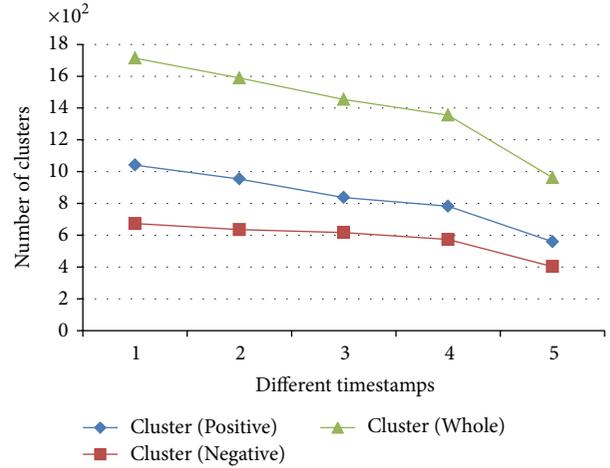


FIGURE 7: The number of congested routes according to the different timestamps.

TABLE 1: The values of parameters.

Parameter	Value
Road network	Oldenburg city
The number of segments	7,035
The number of nodes	6,105
The number of vehicles	10,000~100,000
Velocity	0~120

NETSCAN and the proposed scheme are evaluated according to the number of the vehicles. For the proposed scheme, the congested routes are evaluated in different directions (positive direction and negative direction) and same direction, respectively. PROPOSED (Positive) and PROPOSED (Negative) mean the number of congested routes of each direction of road segments, and PROPOSED (Whole) means the number of congested routes of bidirectional road segments according to the number of vehicles. The results show that the number of the congested routes is increased when the number of the vehicles increases. The number of the congested routes of the proposed scheme is similar when the number of the vehicles is between 20,000 and 30,000. This is because the saturation of each road segment is considered in the proposed scheme. When the width and length of a road segment are large, 20,000 and 30,000 vehicles are not large for the road. Therefore, most of the road segments are not identified as congested routes at first. For NETSCAN scheme, the number of congested routes is increased proportionally with the increase of the number of vehicles.

Figure 7 shows the number of congested routes according to the various timestamps when the time interval is set to 1 hour. The congested routes are evaluated in different directions (positive direction and negative direction) and same direction, respectively. Cluster (Positive) and Cluster (Negative) mean the number of congested routes of each direction of road segments according to the different timestamps. We can see that the number of the congested routes of the positive direction is larger than that of the negative

direction. And the number of the congested routes of the proposed scheme that does not consider the direction of the road segment is larger than that of the congested routes of the proposed scheme considering the direction of the road segment.

## 5. Conclusion

In this paper, we have proposed a congested routes discovering scheme in real road networks. The proposed scheme divides the road into segments with different widths and lengths. It extracts the congested road segments based on the average speed of the vehicles and the saturation degree of a road segment. The final congested routes are computed by performing clustering scheme. The experimental results have shown that the proposed scheme can discover the congested routes in different directions over the existing schemes. In the future, we will show the superiority of our approach by using the real trajectory data of vehicles.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) Support Program (NIPA-2014-H0301-14-1022) supervised by the NIPA (National IT Industry Promotion Agency) and the C-ITRC (Convergence Information Technology Research Center) Support Program (NIPA-2014-H0401-14-1007) supervised by the NIPA (National IT Industry Promotion Agency), by the ICT R&D Program of Discovery Platform for Large-Scale Real Time Data Analysis, MSIP/IITP (14-824-09-001, Development of High Performance Visual Big Data Discovery Platform for Large-Scale Realtime Data Analysis), and by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIP) (no. 2013RIA2A2A01015710).

## References

- [1] Z. Chen, H. T. Shen, and X. Zhou, "Discovering popular routes from trajectories," in *Proceedings of the IEEE 27th International Conference on Data Engineering (ICDE '11)*, pp. 900–911, Hanover, Germany, April 2011.
- [2] P. Cudre-Mauroux, E. Wu, and S. Madden, "TrajStore: an adaptive storage system for very large trajectory data sets," in *Proceedings of the 26th IEEE International Conference on Data Engineering (ICDE '10)*, pp. 109–120, Long Beach, Calif, USA, March 2010.
- [3] E. Kanoulas, Y. Du, T. Xia, and D. Zhang, "Finding fastest paths on a road network with speed patterns," in *Proceedings of the 22nd International Conference on Data Engineering*, p. 10, April 2006.
- [4] G. Gidófalvi, C. Borgelt, M. Kaul, and T. B. Pedersen, "Frequent route based continuous moving object location- and density prediction on road networks," in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '11)*, pp. 381–384, November 2011.
- [5] J.-W. Chang, M.-S. Song, and J.-H. Um, "TMN-tree: new trajectory index structure for moving objects in spatial networks," in *Proceedings of the IEEE 10th International Conference on Computer and Information Technology (CIT '10)*, pp. 1633–1638, Bradford, UK, June–July 2010.
- [6] M. Huang, P. Hu, and L. Xia, "A grid based trajectory indexing method for moving objects on fixed network," in *Proceedings of the 18th International Conference on Geoinformatics*, June 2010.
- [7] J.-I. Won, S.-W. Kim, J.-H. Baek, and J. Lee, "Trajectory clustering in road network environment," in *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM '09)*, pp. 299–305, April 2009.
- [8] G.-P. Roh, J.-W. Roh, S.-W. Hwang, and B.-K. Yi, "Supporting pattern-matching queries over trajectories on road networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 11, pp. 1753–1758, 2011.
- [9] N. Pelekis, I. Kopanakis, E. E. Kotsifakos, E. Frentzos, and Y. Theodoridis, "Clustering trajectories of moving objects in an uncertain world," in *Proceedings of the IEEE International Conference on Data Mining (ICDM '09)*, pp. 417–427, December 2009.
- [10] M. Kimura, S. Inoue, Y. Kakuda, and T. Dohi, "A route discovery method for alleviating traffic congestion based on VANETs in urban transportations considering a relation between vehicle density and average velocity," in *Proceedings of International Symposium on 10th International Symposium on Autonomous Decentralized Systems (ISADS '11)*, pp. 299–302, March 2011.
- [11] L.-Y. Wei, W.-C. Peng, and W.-C. Lee, "Exploring pattern-aware travel routes for trajectory search," *ACM Transactions on Intelligent Systems and Technology*, vol. 4, no. 3, article 48, 2013.
- [12] Y. Xu and J. Wang, "Optimal path solution of urban traffic road," in *Proceedings of the 7th International Conference on Natural Computation (ICNC '11)*, pp. 799–802, July 2011.
- [13] A. Kharrat, K. Zeitouni, I. Sandu-Popa, and S. Faiz, "Characterizing traffic density and its evolution through moving object trajectories," in *Proceedings of the 5th International Conference on Signal Image Technology and Internet Based Systems (SITIS '09)*, pp. 257–263, December 2009.
- [14] K. W. Min, J. W. Kim, and J. H. Park, "Optimal route determination technology based on trajectory querying moving object database," in *Proceedings of International Conference on Database and Expert Systems Applications*, pp. 666–675, 2006.
- [15] V. Tyagi, S. Kalyanaraman, and R. Krishnapuram, "Vehicular traffic density state estimation based on cumulative road acoustics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1156–1166, 2012.
- [16] X. Li, J. Han, J. Lee, and H. Gonzalez, "Traffic density-based discovery of hot routes in road networks," in *Proceedings of SIAM International Conference on Data Mining*, pp. 441–459, 2007.
- [17] H. M. O. Mokhtar, O. Ossama, and M. E. Sharkawi, "A time parameterized technique for clustering moving object trajectories," *Journal of Data Mining and Knowledge Management Process*, vol. 1, no. 1, pp. 14–30, 2011.

- [18] I. Abraham, D. Delling, A. V. Goldberg, and R. F. Werneck, "Alternative routes in road networks," *Journal of Experimental Algorithmics*, vol. 18, article 1.3, 2013.
- [19] A. Selamat, M. Zolfpour-Arokhlo, S. Z. Hashim, and M. H. Selamat, "A fast path planning algorithm for route guidance system," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC '11)*, pp. 2773–2778, Anchorage, Alaska, USA, October 2011.
- [20] H. Gonzalez, J. Han, X. Li, M. Myslinska, and J. P. Sondag, "Adaptive fastest path computation on a road network: a traffic mining approach," in *Proceedings of International Conference on Very Large Data Bases*, pp. 794–805, 2007.
- [21] T. Brinkhoff, "A framework for generating network-based moving objects," *GeoInformatica*, vol. 6, no. 2, pp. 153–180, 2002.

