

DESIGN AND IMPLEMENTATION OF TRAJECTORY DATA MANAGEMENT AND ANALYSIS TECHNOLOGY FRAMEWORK BASED ON SPATIOTEMPORAL GRID MODEL

J. Li¹, J. Q. Liu^{1*}, X. L. Mei¹, W. T. Sun¹, Q. Huang², Y. Y. Zhang¹, L. W. Qiao¹, C. Y. Zhang¹

¹ College of Geoscience and Surveying Engineering, China University of Mining and Technology, Beijing 100083, China – junli@cumtb.edu.cn, (juq.liu, xl.mei, wt.sun, yy.zhang, lw.qiao)@student.cumtb.edu.cn, czhang@cumtb.edu.cn

² Service Lab, Huawei Research Institute, Huawei Technology Co. Ltd, Beijing 100095, China – huangqian16@huawei.com

KEY WORDS: Trajectory Representation, Trajectory Data Management, Trajectory Analysis, Spatiotemporal Grid Model

ABSTRACT:

The trajectory data generated by various position-aware devices is widely used in various fields of society, but its conventional vector representation and various analysis algorithms based on it have high computational complexity. This makes it difficult to meet the application requirements of real-time or near real-time management and analysis of large-scale trajectory data. In view of the above challenges, this paper proposes a trajectory data management and analysis technology framework based on the Spatiotemporal Grid Model (STGM). First, the trajectory data is represented by spatiotemporal grid encoding instead of vector coordinates, and it can achieve dimensionality reduction and integrated management of high-dimensional heterogeneous trajectory data. Second, the trajectory computing and analysis methods based on STGM are introduced, which reduce the computing complexity of algorithms. Furthermore, various types of trajectory mining and applications are realized on the basis of high-performance computing technologies. Finally, a trajectory data management and analysis prototype system based on the STGM is developed, and experimental results verify the reliability and effectiveness of the proposed technology framework.

1. INTRODUCTION

The rapid development of mobile internet technology has spawned a large amount of mobile trajectory data (Wu et al., 2019). These trajectory data are widely used in smart transportation (Li et al., 2012), urban computing (Zheng et al., 2015), social sensing (Liu et al., 2015) and other fields because of their rich spatiotemporal location and semantic information. For example, Li et al. (2019) used the vehicle trajectory data to extract coach operation information such as coach stations, routes and timetables, which provided data support for China's national road passenger transportation ticketing platform. Some studies used taxi trajectory data to conduct passenger-finding strategies, spatiotemporal analysis of public transportation, road networks update and other studies to optimize urban traffic (Wu et al., 2016; Tang et al., 2017; Tu et al., 2018). Scholars also used mobile phone traces to study residents' mobility laws to assist scientific and smart city planning (Jiang et al., 2013; Chen et al., 2018). On the other hand, the huge trajectory data also bring challenges to data management and analysis due to its characteristics of large-scale, dynamic update, multi-source heterogeneity and high-dimensional (Feng, Zhu, 2016; Li et al., 2016). The two typical representation models (vector and raster model) are difficult to cope with those problems and cannot satisfy the real-time or near real-time trajectory data mining and application needs.

In recent years, the rapid development of computer technology has promoted rapid evolution of Discrete Global Grid System (DGGS). The characteristics of discreteness, multi-level, and low-dimensional of DGGS provide a new research perspective

for efficient management and analysis of massive trajectory data (Zhou et al., 2009; Goodchild, 2018). Specifically, the discreteness of DGGS not only meets the requirements of computer for discretizing storage, but also facilitates the distributed processing of massive spatiotemporal data. The multi-level grid models can adaptively use the grid code to calculate and analyze the problems at different scales to improve efficiency. The low-dimensional encoding provides a basis for efficient and flexible storage and organization of trajectory data (Chen et al., 2002; Purss et al., 2016).

2. THE FRAMEWORK DESIGN

The trajectory data management and analysis technology framework based on the Spatiotemporal Grid Model (STGM) mainly includes five parts: multi-source trajectory data, spatiotemporal grid model, trajectory computing and analysis methods, high-performance computing (HPC) and trajectory mining applications, as illustrated in Figure 1, which provides the solutions for the knowledge discovery of massive trajectory data and various applications. The STGM represents multi-source trajectory data through spatiotemporal grid encoding to achieve data fusion and dimensionality reduction. On this basis, the common trajectory computing and analysis methods are transformed based on the low-complexity code operations to accelerate trajectory mining. The high-performance computing technologies provide distributed storage resources and concurrent computing resources. On the top are various types of trajectory mining and applications.

* Corresponding author

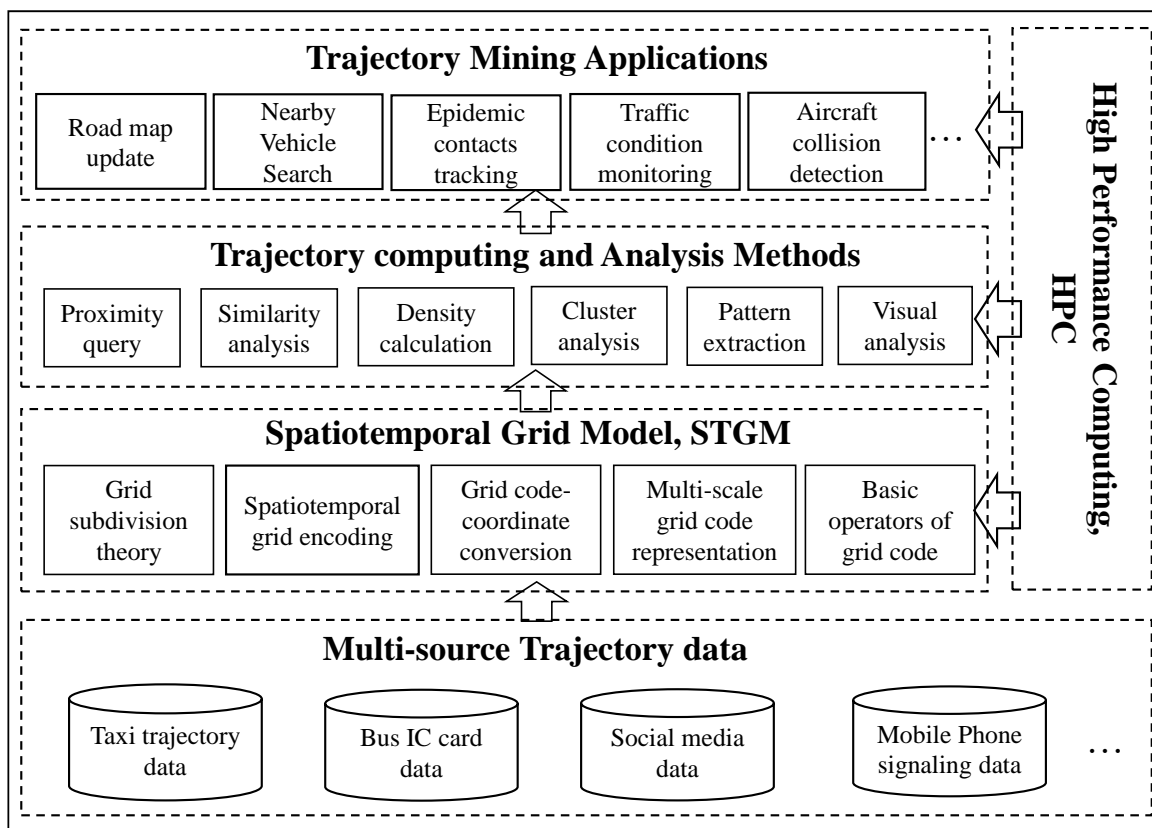


Figure 1. Technology framework of trajectory data management and analysis based on the STGM

2.1 Spatiotemporal Grid Model

The STGM uses the grid subdivision theory and spatial grid encoding technology to replace the vector floating-point coordinates with local unit address codes (Cui et al., 2007; Sun et al., 2008; Wan, Cao, 2016; Qian et al., 2019; Guo et al., 2019) (Figure 2a). Then, the time dimension is taken into account to realize the one-dimensional encoding representation of space-time information. Furthermore, the trajectory data is mapped to the spatiotemporal grids to realize multi-scale coding

representation (Figure 2b). In this way, the dimensionality reduction of the high-dimensional trajectory data is achieved, which greatly reduces the complexity of trajectory data storage, management and analysis. What's more, the STGM develops the conversion function between grid code and coordinates. Finally, two basic operators—proximity grid code query and spatial geometric measurement are developed, and they provide the foundation for upper-layer trajectory analysis algorithm.

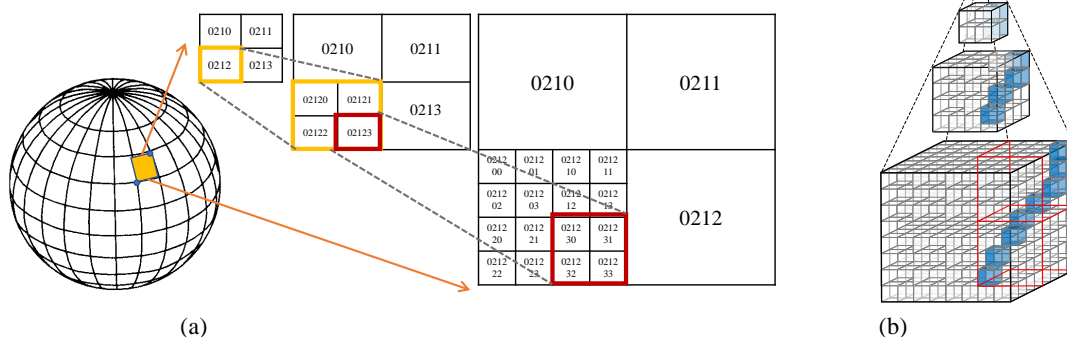


Figure 2. Spatiotemporal grid model, (a) Spatial grid coding, (b) Multi-scale coding representation of trajectory

2.2 Trajectory Computing and Analysis Methods

The change of the trajectory data representation model will inevitably lead to reformation of the upper-level trajectory analysis algorithms. It is necessary to fully make use of the trajectory grid code representation characteristics to reform or improve a set of computing and analysis methods based mainly on the following strategies: a) The fast query of trajectories is realized based on the structural characteristics of "grid-coding is indexing, indexing is grid-coding", such as spatiotemporal proximity query and road matching with trajectory points. The basic idea is to organize the trajectory codes as a data index, so the trajectory query can be implemented on the index layer, which greatly accelerates the query operation of trajectories. In addition, instructed by the principle that the code prefixes of trajectories in the same grid are close in time and space, the fast spatiotemporal query can be achieved using code matching. b) The trajectory analysis algorithms can be transformed by a set of low complexity of code operations, such as trajectory distance measurement. The classical trajectory distance measurement algorithms include Dynamic Time Warping (DTW) (Keogh et al., 2000), Longest Common Subsequence (LCSS) (Vlachos et al., 2002), Edit Distance on Real sequence (EDR) (Chen et al., 2005), Fréchet distance (Fréchet et al., 1906) and Hausdorff distance (Lee et al., 2007). The time complexity of these methods is $O(n*m)$ (n and m are the number of points of the two trajectories respectively), which makes the calculation very time-consuming. Nevertheless, the trajectory distance based on grid code can be calculated through the Jaccard distance, Simpson coefficient, dice coefficient etc., which are very fast calculation operations. On the basis of this, most other trajectory analysis algorithms can be transformed, simplified and accelerated, such as trajectory similarity analysis, trajectory clustering. c) The flexibility of trajectory computing and analysis can be realized by the multi-scale grid code representation of trajectory. Although the classical vector coordinates have a high accuracy, it is difficult to handle cross-scale analysis tasks such as a multi-mode traffic travel analysis, including large-scale flight trajectories, medium-scale intra-city and inner-city trajectories, and walking trajectories at a small scale. Based on the grid code, the cross-scale analysis can be easily achieved by selecting the trajectory codes at suitable scales. d) The multi-scale characteristics of trajectory codes can also serve for visual analysis. The trajectory data at a corresponding scale is visualized according to the scale of web view range to accelerate the visual analysis, which is similar to the image pyramid structure.

2.3 High Performance Computing

In the era of big data, the mining of massive trajectory data often requires the support of high-performance computing frameworks such as parallel computing and distributed computing (Gao et al., 2017). The STGM's discrete characteristic can be well combined with high performance computing, which realizes the distributed storage and concurrent computing of big trajectory data, and provides bidirectional power for the management and analysis of trajectory data. The specific performance includes: First, the trajectory codes are discretized by slices (time slices), layers (different grid-coding levels), and blocks (space cells), and stored in existing distributed storage systems, such as MongoDB, HBase, PostgreSQL and other distributed databases,

to achieve storage and management of massive trajectories. Then, the trajectory data is automatically distributed to each node under the distributed computing framework (such as Hadoop, Spark) or GPU parallel computing frameworks. Finally, the concurrent query and computing of massive trajectory data is implemented.

2.4 Trajectory Mining Applications

Trajectory mining applications often require frequent or real-time services with high time efficiency, such as regular updates of road maps, real-time search of nearby vehicles, rapid filtering of epidemic contacts, real-time monitoring of traffic conditions, aircraft collision detection. The proposed trajectory data management and analysis technology framework based on the STGM uses the advantages of trajectory grid coding to accelerate trajectory computing and analysis, which makes it possible to meet these real-time or near real-time application requirements. For example, (a) for the real-time dispatch of massive taxis, the proposed technology framework can quickly gather people's travel demand, simultaneously conduct real-time query of nearby taxis, and provide solutions for intelligent dispatch of taxis and peak hour pricing service. (b) For the filtering of epidemic contacts, we can take a large-scale individual trajectories as the target, and conduct similar trajectory analysis based on STGM to quickly identify and track suspected epidemic contacts among a large number of people to block the virus transmission chain. (c) In order to solve the problem of collision detection among flying aircrafts in a large-scale, real-time scenario, the flying trajectories of aircrafts are grid-coded based on STGM. Then, the inclusion judgment based on grid codes is used to improve the efficiency of aircraft collision detection and ensure the flight safety of aircrafts (Zheng et al., 2019).

3. EXAMPLE OF APPLICATION

Based on the above theory and technology, a trajectory data management and analysis prototype system based on the STGM was developed (Figure 3). The system manages all the taxi trajectory data of Beijing, China, a total of approximately 70,000 taxis, generating approximately a dozen GB of trajectory data every day.

The trajectory analysis functions include spatiotemporal proximity query of trajectory, top-k similar trajectory query and other functions. The spatiotemporal proximity query of trajectory outputs adjacent trajectories of a target trajectory within a specific radius of a certain space-time position (Figure 4a). Experimental results validate that the proximity query under tens of millions of trajectory points can be finished in less than a second with the grid index. Top-K similar trajectory query is to get the most K similar trajectories of a target trajectory in a large-scale dataset (Figure 4b). The system realizes fast query of similar trajectories through multi-scale code and highly efficient code operations. Experimental results show that the computing time is shortened by about two orders of magnitude with a similar accuracy, compared with the classic trajectory similarity analysis algorithms (e.g. DTW, LCSS, EDR). The multi-scale trajectory density calculation module calculates the trajectory density in each grid, and realizes fast switching among different view ranges (Figure 4c).

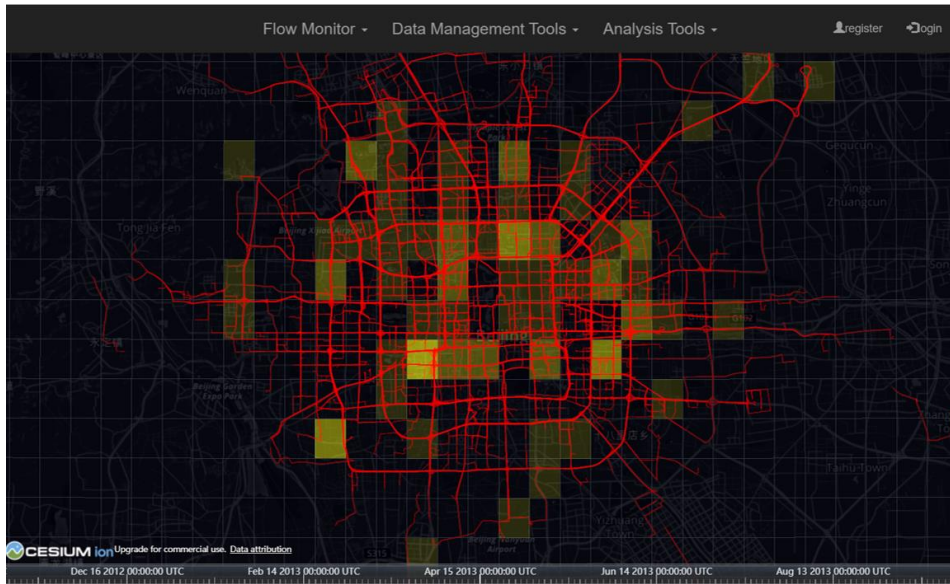
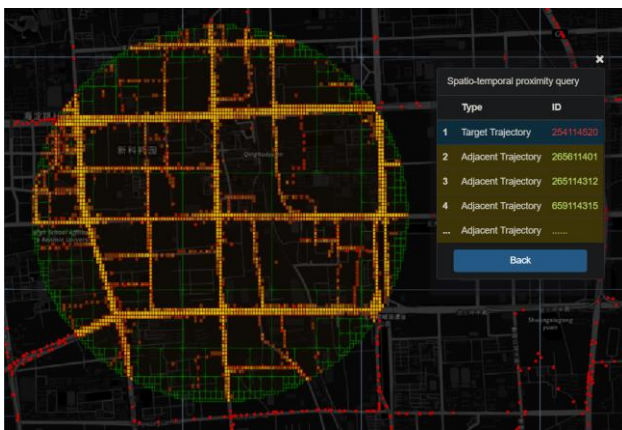
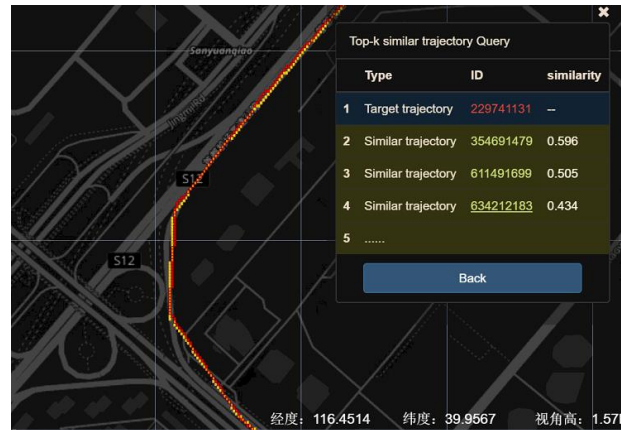


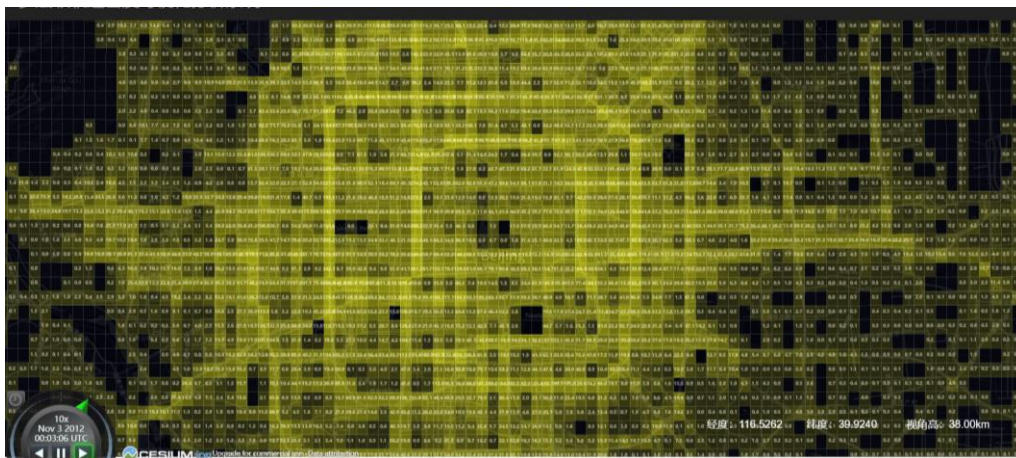
Figure 3. Interface of trajectory data management and analysis prototype system based on STGM



(a)



(b)



(c)

Figure 4. Examples of trajectory analysis tools in the system, (a) Spatiotemporal proximity query of trajectory, (b) Top-k similar trajectory query, (c) Fast calculation of multi-scale trajectory density

4. CONCLUSION

This paper proposes a trajectory data management and analysis technology framework based on the STGM, which mainly includes five parts: heterogeneous multi-source trajectory data, STGM, HPC, trajectory computing and analysis methods, and trajectory mining applications. In addition, we analyze and summarize the advantages and characteristics of storing trajectory data and simplifying the analysis algorithms based on spatiotemporal grid-coding. The computing and analysis algorithms of trajectory data are accelerated by using the features of grid index, multi-scale coding, and transformed analysis algorithms, which provide the possibility for real-time or near real-time trajectory mining applications. Finally, we developed a trajectory data management and analysis prototype system based on the STGM, which verified the effectiveness of the theories and technical methods presented in the paper. This work provides technical support for the management and analysis of massive trajectory data in future, and better serves various fields such as smart transportation, urban computing, city planning and social sensing.

ACKNOWLEDGEMENTS

This research is financially supported by National Natural Science Foundation of China (No. 41971355), the Open Project of State Key Laboratory of Resources and Environmental Information System, Huawei Corporation and the Yueqi Young Scholars Program of China University of Mining and Technology at Beijing.

REFERENCES

- Chen, B.Y., Wang, Y., Wang, D., Li, Q., Lam, W.H., Shaw, S.L., 2018: Understanding the impacts of human mobility on accessibility using massive mobile phone tracking data. *Annals of the American Association of Geographers*, 108(4), 1115-1133.
- Chen, L., Özsu, M.T., Oria, V., 2005. Robust and fast similarity search for moving object trajectories. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, 491-502.
- Chen, S. P., Zhou, C.H., Chen, Q.X., 2002: Grid mapping and grid computing. *Science of Surveying and Mapping* 27, no. 4, 1-6.
- Cui, M.J., Zhao, X.S., 2007: Tessellation and distortion analysis based on spherical DQG. *Geography and Geo-Information Science*, 23(6), 23-25.
- Feng, Z., Zhu, Y., 2016: A survey on trajectory data mining: Techniques and applications. *IEEE Access*, 4, 2056-2067.
- Fréchet, M.M., 1906: Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 22(1), 1-72.
- Gao, Q., Zhang, F.L., Wang, R.J., Zhou, F., 2017. Trajectory big data: a review of key technologies in data processing. *J. Softw.*, 28(4), 959-992.
- Goodchild, M.F., 2018: Reimagining the history of GIS. *Annals of GIS*, 24(1), 1-8.
- Guo, N., Xiong, W., Wu, Y., Chen, L., Jing, N., 2019: A geographic meshing and coding method based on adaptive Hilbert-Geohash. *IEEE Access*, 7, 39815-39825.
- Jiang, S., Fiore, G.A., Yang, Y., Ferreira Jr, J., Frazzoli, E., González, M.C., 2013. A review of urban computing for mobile phone traces: current methods, challenges and opportunities. In *Proceedings of the 2nd ACM SIGKDD international workshop on Urban Computing* (pp. 1-9).
- Keogh, E.J., Pazzani, M.J., 2000. Scaling up dynamic time warping for datamining applications. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 285-289.
- Lee, J.G., Han, J., Whang, K.Y., 2007. Trajectory clustering: a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, 593-604.
- Li, J., Li, Q., Zhu, Y., Ma, Y., Xu, Y., Xie, C., 2019: An Automatic Extraction Method of Coach Operation Information from Historical Trajectory Data. *Journal of Advanced Transportation*, 2019, 2019(P1):449-463.
- Li, J., Qin, Q., Xie, C., Zhao, Y., 2012: Integrated use of spatial and semantic relationships for extracting road networks from floating car data. *International Journal of Applied Earth Observation and Geoinformation*, 19, 238-247.
- Li, J., Wang, J., Zhang, J., Qin, Q., Jindal, T., Han, J., 2016: A probabilistic approach to detect mixed periodic patterns from moving object data. *Geoinformatica*, 20(4), 715-739.
- Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G.H., Shi, L., 2015: Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105(3), 512-530.
- Purss, M.B., Gibb, R., Samavati, F., Peterson, P., Ben, J., 2016. The ogc® discrete global grid system core standard: A framework for rapid geospatial integration. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 3610-3613).
- Qian, C., Yi, C., Cheng, C., Pu, G., Wei, X., Zhang, H., 2019: Geosot-based spatiotemporal index of massive trajectory data. *ISPRS International Journal of Geo-Information*, 8(6), 284.
- Sun, W., Cui, M., Zhao, X., Gao, Y., 2008. A global discrete grid modeling method based on the spherical degenerate quadtree. In *2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing. IEEE*, (Vol. 2, pp. 308-311).
- Tang, L., Sun, F., Kan, Z., Ren, C., Cheng, L., 2017: Uncovering distribution patterns of high performance taxis from big trace data. *ISPRS International Journal of Geo-Information*, 6(5), 134.
- Tu, W., Cao, R., Yue, Y., Zhou, B., Li, Q., Li, Q., 2018: Spatial variations in urban public ridership derived from GPS trajectories and smart card data. *Journal of Transport Geography*, 69, 45-57.

Vlachos, M., Kollios, G., Gunopoulos, D., 2002. Discovering similar multidimensional trajectories. In *Proceedings 18th international conference on data engineering. IEEE*, 673-684.

Wan, G., Cao, X.F., 2016: The Historical Evolution and Reflection of Geospatial Information Grid. *Acta Geodaetica et Cartographica Sinica*, 45(S1), 15-22

Wu, H.Y., Huang, R., You, L., Xiang, L.G., 2019: Recent Progress in Taxi Trajectory Data Mining. *Acta Geodaetica et Cartographica Sinica*, 48(11), 1341-1356.

Wu, T., Xiang, L., Gong, J., 2016: Updating road networks by local renewal from GPS trajectories. *ISPRS International Journal of Geo-Information*, 5(9), 163.

Zheng, Y., 2015: Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3), 1-41.

Zheng, Y., Liu, J., Li, J., Xu, Y., Huang, Q., Zhu, Y., Pei, Y., 2019. Design of Fine Management System for Civil Aviation Airspace Resources Based on Spatiotemporal Grid Model. In *2019 IEEE 1st International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, 547-551.

Zhou, C.H., Ou, Y., Ma, T., 2009: Progresses of geographical grid systems researches. *Prog. Geogr.*, 28, 657-662.