Visualization aware trajectory sampling for urban traffic data

Category: Research

Paper Type: please specify

Fig. 1. In the Clouds: Vancouver from Cypress Mountain. Note that the teaser may not be wider than the abstract block.

Abstract—Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

Index Terms—Radiosity, global illumination, constant time

1 Introduction

Nowadays, the widely used location-acquisition devices lead to an explosive increase of the movement data which is recorded in the form of trajectories. For example, the taxis trajectory is one of the common studied movement data which is always considered as the representative of human movement trace in a city. Using the taxi dataset in Shenzhen as an example, more than **(size) of trajectory data can be collected every day, which records (distance) by sampling locations. The analysis over these databases can be applied in many fields such as traffic management [15], urban planning, route recommendation [20] and location-based services [5, 19].

Visualizing trajectories is a challenging task. The most popular and conventional method is the line-based visualization [3]: connecting the passing points of movement objects by polylines. The current visualization tools always don't scale well for the presentation of very large trajectory dataset due to the two challenges, visual clutter and limited rendering speed, which hinders the abilities of human-users for interactively exploring the dataset and identifying the movement patterns. In recent years, most of the visualization research works mainly try to address the visual clutter issue by proposing new techniques such as the spatial aggregation [14, 17], edge bundling [13, 18] and density map [4, 11]. Instead, in this paper, we focus on the challenge of inefficient rendering in the large trajectory dataset by involving data sampling techniques.

Using ** dataset as an example, figure 2 demonstrates the rendering time at each dataset size, which shows that normal method takes more than ** minutes to generate the visualization, which is far beyond the human-acceptable response time for the interactive exploration [12].

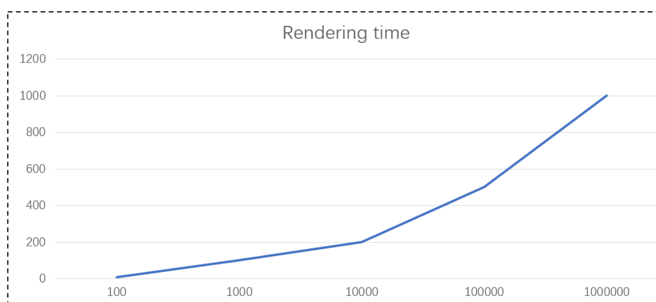

Fig. 2. x-data size; y-rendering time

To handle the big dataset, many visualization products such as Spotfire [] and Tableau [] support advanced database management systems as a "backend" for the efficient data processing the query. One work closely related to ours is ScalaR [1], which adds a reduction layer between visualization layer and data management layer. The reduction layer uses an uniform random sampling method to sample data once the query results are large enough, thus to reduce the amount of data to be visualized. Further more, Park et al. propose VAS [7] which implements new sampling techniques to guarantee the visual quality. However, these sampling techniques are designed for the simple dataset, and have been approved effective in scatter plot or map plot. However, the trajectory sampling is more challenge due to the complexity of

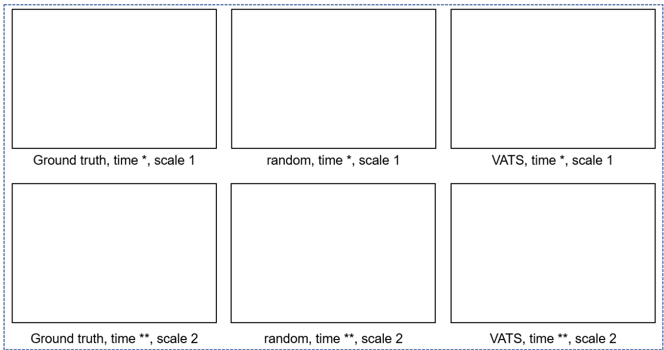

Fig. 3. three columns (ground truth, random sampling, proposed method), two rows(top level, middle level)

data form(e.g. varying lengths, lack of compact representation, difficulty in measuring the similarity) that makes traditional density-biased sampling techniques inappropriate.

In our method, we extend the motivation of visualization-aware sampling to trajectory dataset. We propose a novel sampling strategy, **v**isualization **a**ware **t**rajectory **s**ampling(VATS), that produces high-visual-quality line-based trajectory visualization at certain degree(arbitrary) zooming resolutions. In this paper, we first proposed the visual fidelity loss function which effectively evaluates the visual loss of the sampling method. Then we minimize the loss function by transforming this problem to an optimization problem. Several solutions for efficiently solving the optimization problem are discussed. Figure 3 depicts an comparison among the ground truth, uniform random sampling and our proposed method. By limiting the sampling set size, the proposed method generates a higher-fidelity visualization and support the multi-resolution very well.

We summarize our contribution as follows:

- We formulate VATS as an optimization problem.
- We prove VAST problem is NP-hard and offer an efficient approximation algorithms.
- We conduct several experiments using real-world data to demonstrate the effectiveness of the proposed method in comparison with random uniform sampling.

The remaining parts are constructed as follows: section 2 discusses the related work. In section 3, we identify the specific problem and provide an overview of our solution. We define the problem and propose the solution in the section 4 and 5. The implementation and experiment setting are introduced in section 6. In section 7, we conduct case studies and user studies to evaluate our approach. Finally, we conclude this paper and propose the possible future directions in section 8.

2 Related Work

The most related techniques to our work include the visual analysis of trajectory dataset, the methodology of large data visualization and data sampling.

2.1 Trajectory analysis

Trajectory, consisting of a sequence of spatial locations, is the most common form of the object movement. To support the understanding and analysis of the trajectory dataset, many visualization and visual analytics system are developed. The detailed summary of these work is presented in [3]. These techniques can be classified into three categories according to visualization form: point-based visualization, line-based visualization and region-based visualization.

The point-based visualization capture the basic spatial distribution of the passing points of the moving object []. Furthermore, many density-based methods such as the kernel density estimation(KDE) are applied based on the point-based visualization, by the sacrifice of the detail the information of trajectories, these methods alleviate the visual clutter caused by large amount of data. Furthermore, to be better applied in the city environment, advanced KDE techniques are developed to capture the moving patterns along the road networks []. In the study of urban traffic, the point-based visualization can capture the hot regions, but unable to identify the movement of the individual case and reveal the moving information such as the direction and origin-destination [3]. Line-based techniques are the most commonly used visualization methods which present the trace of the movement as polylines, thus to preserve the continuous moving information []. However, due the large amount of the trajectories, the line-based methods always cause serious visual clutter due to the cross of the polylines []. To alleviate this problem, the clustering techniques are applied in the preprocessing step before visualization and then visualize the clusters by ribbon [], glyph [] or sankey diagrams []. Moreover, advanced interaction techniques [], sampling techniques [] and edge bundling techniques [18] are also developed to better present the movement patterns. The region based techniques divide the whole region into sub-regions in advance and then visualize the traffic situation before the sub-regions. These methods visualize the macro-pattern very well by leveraging different aggregation techniques such as the administrative regions [], uniform grid [] and spatial clustering results [].

2.2 Large data visualization

Another topic related to our work is the large data visualization. The movement dataset, such as the urban traffic, always contains millions to billions of trajectories, which is challenge to analyze and visualize. Three methods are applied to support the visual analytics of large datasaet: interactive method, aggregation method and reduction method. All these methods aim to reduce the amount of the rendering items to support the efficient visualization.

More discussion about the limitations

2.3 Data sampling techniques

As we have discussed in the previous sections, the sampling methods reduce the data volume by directly selecting the representative items. Current advancing sampling techniques in the visualization domain are mostly designed for the scatter plot and aim to not only solve the overdrawing of the points but also try to preserve the information distribution of the data items. Some works design advanced sampling algorithms to preserve the meaningful data items according to the analyzing requirement such as the mulit-class data analysis [2] and hierarchical exploration []. Furthermore, to the usage of more visual channels of the points other than location such as color [2], size [16] and opacity are discussed. Closely related to our work, Park et al. [7] proposed the visualization-aware techniques for the scatter plot. They proposed visualization-inspired loss which effectively evaluates the visual loss of the sampling result and validates the proposed method based on three common visualization goals: regression, density estimation and clustering.

In comparison with the sampling techniques for scatter plot, the trajectory sampling is more challenging because of the complexity of the trajectories [10]. Most of the existing trajectory sampling techniques cluster the trajectories first [8] and then select the most representative trajectories from each cluster, which highly depend on the distance calculation [9] and clustering algorithms. Some techniques further focus on the clustering and sampling of trajectory segments instead of the whole trajectories [6]. None of these sampling techniques are proposed from the motivation visualization-aware...

3 Problem Statement

3.1 Problem Identification

- Definition of visual quality
- Evaluate visual quality

3.2 VATS Overview

4 Problem Formulation

5 Problem Solving

6 Implementation

Platform/language/space-time consume

7 Evaluation

7.1 Case studies

- Data description
- Anomaly case
- Visual quality case

7.2 User study

- Visual similarity
- Identify outliers
- Trustiness

7.3 Expert overview

8 Discussion and future work

References

[1] L. Battle, M. Stonebraker, and R. Chang. Dynamic reduction of query result sets for interactive visualizaton. In *2013 IEEE International Conference on Big Data*, pp. 1–8. IEEE, 2013.

[2] H. Chen, W. Chen, H. Mei, Z. Liu, K. Zhou, W. Chen, W. Gu, and K.-L. Ma. Visual abstraction and exploration of multi-class scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1683–1692, 2014.

[3] W. Chen, F. Guo, and F.-Y. Wang. A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems*, 16(6):2970–2984, 2015.

[4] O. D. Lampe and H. Hauser. Interactive visualization of streaming data with kernel density estimation. In *2011 IEEE Pacific visualization symposium*, pp. 171–178. IEEE, 2011.

[5] D. Liu, D. Weng, Y. Li, J. Bao, Y. Zheng, H. Qu, and Y. Wu. Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations. *IEEE transactions on visualization and computer graphics*, 23(1):1–10, 2016.

[6] C. Panagiotakis, N. Pelekis, I. Kopanakis, E. Ramasso, and Y. Theodoridis. Segmentation and sampling of moving object trajectories based on representativeness. *IEEE Transactions on Knowledge and Data Engineering*, 24(7):1328–1343, 2011.

[7] Y. Park, M. Cafarella, and B. Mozafari. Visualization-aware sampling for very large databases. In *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, pp. 755–766. IEEE, 2016.

[8] N. Pelekis, I. Kopanakis, E. Kotsifakos, E. Frentzos, and Y. Theodoridis. Clustering trajectories of moving objects in an uncertain world. In *2009 Ninth IEEE international conference on data mining*, pp. 417–427. IEEE, 2009.

[9] N. Pelekis, I. Kopanakis, G. Marketos, I. Ntoutsi, G. Andrienko, and Y. Theodoridis. Similarity search in trajectory databases. In *14th International Symposium on Temporal Representation and Reasoning (TIME'07)*, pp. 129–140. IEEE, 2007.

[10] N. Pelekis, I. Kopanakis, C. Panagiotakis, and Y. Theodoridis. Unsupervised trajectory sampling. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 17–33. Springer, 2010.

[11] R. Scheepens, N. Willems, H. van de Wetering, and J. J. van Wijk. Interactive visualization of multivariate trajectory data with density maps. In *2011 IEEE Pacific Visualization Symposium*, pp. 147–154. IEEE, 2011.

[12] B. Shneiderman. Response time and display rate in human performance with computers. *ACM Computing Surveys (CSUR)*, 16(3):265–285, 1984.

[13] M. Thöny and R. Pajarola. Vector map constrained path bundling in 3d environments. In *Proceedings of the 6th ACM SIGSPATIAL International Workshop on GeoStreaming*, pp. 33–42, 2015.

[14] T. Von Landesberger, F. Brodkorb, P. Roskosch, N. Andrienko, G. Andrienko, and A. Kerren. Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. *IEEE transactions on visualization and computer graphics*, 22(1):11–20, 2015.

[15] Z. Wang, T. Ye, M. Lu, X. Yuan, H. Qu, J. Yuan, and Q. Wu. Visual exploration of sparse traffic trajectory data. *IEEE transactions on visualization and computer graphics*, 20(12):1813–1822, 2014.

[16] A. Woodruff, J. Landay, and M. Stonebraker. Constant density visualizations of non-uniform distributions of data. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, pp. 19–28, 1998.

[17] W. Zeng, C.-W. Fu, S. M. Arisona, and H. Qu. Visualizing interchange patterns in massive movement data. In *Computer Graphics Forum*, vol. 32, pp. 271–280. Wiley Online Library, 2013.

[18] W. Zeng, Q. Shen, Y. Jiang, and A. Telea. Route-aware edge bundling for visualizing origin-destination trails in urban traffic. In *Computer Graphics Forum*, vol. 38, pp. 581–593. Wiley Online Library, 2019.

[19] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Collaborative location and activity recommendations with gps history data. In *Proceedings of the 19th international conference on World wide web*, pp. 1029–1038, 2010.

[20] Y. Zheng and X. Xie. Learning travel recommendations from user-generated gps traces. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1):1–29, 2011.