SIGIR 2021 Submission 321

Submission information updates are disabled. If you need to update your paper, contact SIGIR 2021 chairs.

For all questions related to processing your submission you should contact the conference organizers. Click here to see information about this conference.

All **reviews sent to you** can be found at the bottom of this page.

Submission 321	
Title:	CheetahTraj: Quality and Efficiency in Large-scale Trajectory Data Visual Exploration
Paper:	(Feb 10, 01:59 GMT)
Track:	Long paper
Author keywords:	trajectory visualization data sampling interactive trajectory exploration
Topics:	Domain-specific applications
Abstract:	Visualizing large-scale trajectory data is a core subroutine for many applications, e.g., traffic management, urban planning, and route recommendation. However, naively visualizing all trajectories for a target region could result in long delay due to large data volume. Ad-hoc sampling can reduce visualization time but may harm visual quality, i.e., generating visualizations that look substantially different from the exact one. In this paper, we propose the CheetahTraj framework to provide high quality trajectory visualization for arbitrary target region with low latency. To this end, we first define a natural pixel-based visual quality function to measure the similarity between two visualizations and formulate a quality optimal trajectory sampling problem. Then, we design the VQGS and VQGS+ algorithms to solve the trajectory sampling problem, which not only provide guaranteed visual quality but also reduce visual clutter. To generate quality guaranteed trajectory samples with high efficiency, we develop quad-tree-based index that allows to use trajectory samples computed offline. Extensive experiments (i.e., case-, user-, and quantitative- studies) are conducted on 3 real world trajectory datasets to verify visualization quality and efficiency of CheetahTraj. The results show that CheetahTraj consistently provide high quality visualizations and its visualization time is orders of magnitude shorter than visualizing all trajectories.
Submitted:	Feb 01, 01:04 GMT
Last update:	Feb 01, 01:04 GMT
I understand that authors cannot be added after paper submission	Yes
I certify that this paper is only submitted to SIGIR 2021	Yes
I understand that I, or one of my co-authors, is obliged to attend the conference once accepted	Yes
I understand that I and all my co-authors will be held accountable	Yes

for any issues with the work	
Student paper	No
Author conflicts:	none

Authors

first name	last name	email	country	affiliation	Web page	corresponding?
Qiaomu	Shen	joyshen06@gmail.com	China	Southern University of Science and Technology		
Chaozu	Zhang	11712021@mail.sustech.edu.cn	China	Southern University of Science and Technology		
Xiao	Yan	yanx@sustech.edu.cn	China	Southern University of Science and Technology		
Chuan	Yang	11612732@mail.sustech.edu.cn	China	Southern University of Science and Technology		
Dan	Zeng	zengd@sustech.edu.cn	China	Southern University of Science and Technology		
Wei	Zeng	wei.zeng@siat.ac.cn	China	Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences		
Bo	Tang	tangb3@sustech.edu.cn	China	Southern University of Science and Technology		✓

Reviews

Review 1

Relevance to SIGIR:	**2**: (Marginally relevant and there's an obviously better venue.)
Quality of Presentation:	**3**: (Missing a few important details but the major points were clear.)
Originality of Work:	**3**: (Somewhat conventional: A number of people could have come up with this if they thought about it for a while.)
Technical Soundness:	**4**: (The authors mostly use appropriate description of technical facts but there are some minor mistakes.)
Impact of Contributions:	**4**: (Some of the ideas, results, or resources will substantially help other people's ongoing research.)
Adequacy of Citations:	**3**: (Just good enough: Review is OK; just about enough past work is covered to make it acceptable.)
Reproducibility of Methods/Findings:	**4**: (Could be reproduced: Researchers could reproduce the methods and results with some effort.)
Overall evaluation:	**-2**: (Reject (the paper has substantial problems relative to possible strengths, and it should be rejected)) The paper focuses on the problem of trajectory visualization, and proposes a visualization framework by sampling-based strategies. Specifically, the authors develop two solutions, i.e., VQGS and VQGS+. Experiments on real-life datasets demonstrate the efficiency and effectiveness. However, many issues are requried to address. D1. The relevance to the topics of SIGIR is not clear. First, this paper mainly studies the trajectory visualization problem with sampling strategies. Please clarify how does this framework interact with users. Second, the proposed framework is established on very small data (i.e., the data size is no more than 1.5GB), which is not‟large-scale” as claimed in the paper. D2. In the Introduction section, the authors state that the existing techniques cannot solve the problem studied in this paper, i.e., low quality or high latency. Please add more explanations. D3. The related work is not sufficient as many related literatures are missing as listed below. [1] Yu et al. Turbocharging Geospatial Visualization Dashboards via a Materialized Sampling Cube Approach. In ICDE, 1165-1176, 2020. [2] Chan et al. QUAD: Quadratic-Bound-based Kernel Density Visualization. In SIGMOD, 35-50, 2020.

	D4. In Section 4, the space and time complexity of the proposed algorithms is not analyzed. D5. The proposed framework has not been experimentally evaluated against existing techniques [1, 2]. In contrast, this paper only conducts experimental comparisons with self-designed methods. Hence, the contributions of this paper are not clear. D6. More details (e.g., the min/max/avg length of trajectories) about the evaluated datasets are needed. D7. The dataset size is too small compared with existing work, especially that the authors claimed they are targets for “large-scale” trajectory data visualization. D8. The paper writing needs carefully revised. -In Section 3.1 (Page 3), where is “Figure 2(A)”? -In Section 7 (Page 9), “including VQGS and VQGS” - There are so many symbols, and an additional symbol table would help to follow.

Review 2

Relevance to SIGIR:	**4**: (Relevant: Interesting to some SIGIR participants.)
Quality of Presentation:	**3**: (Missing a few important details but the major points were clear.)
Originality of Work:	**3**: (Somewhat conventional: A number of people could have come up with this if they thought about it for a while.)
Technical Soundness:	**4**: (The authors mostly use appropriate description of technical facts but there are some minor mistakes.)
Impact of Contributions:	**4**: (Some of the ideas, results, or resources will substantially help other people's ongoing research.)
Adequacy of Citations:	**2**: (Not good enough: Authors have written a review, but its coverage misses many important past papers.) The following papers studying sampling on POIs for visualisation should be carefully reviewed as well: [1] Efficient Selection of Geospatial Data on Maps for Interactive and Visualized Exploration [2] An efficient sampling method for characterizing points of interests on maps Visual constraint is considered in: [3] Efficient spatial sampling of large geographical tables [4] Declarative cartography: In-database map generalization of geospatial datasets [5] Multiresolution select-distinct queries on large geographic point sets [6] Viewing streaming spatially-referenced data at interactive rates
Reproducibility of Methods/Findings:	**5**: (Easy to reproduce: Researchers could easily reproduce the methods and results described in the paper without much difficulty.)
Overall evaluation:	**-1**: (Weak reject (the paper is near the borderline, but on balance the problems outweigh the strengths, so it should not be accepted)) This paper solves a trajectory visualisation problem by introducing a novel QOSP sampling problem. The authors proposed VQGS which is a naive greedy solution with lazy update optimisation. Then VQGS+ is proposed by introducing a theta parameter to ignore some neighbour trajectories. To optimise the performance further, authors propose CheetahTraj to precompute offline for VQGS+. Strong points: 1) The problem definition is sound enough to illustrate the visualisation problem. 2) Empirical studies show the proposed algorithms are efficient. Weak points: 1) Lack of theatrical support: a) The NP-hard proof in Sec 3.2 is not complete. To prove a problem is NP-hard, reducing it to another NP-hard problem is a common way. However, a complete proof need to reduce the NP-hard problem to QOSP problem by giving an instance. b) Both VQGS and VQGS+ should have a time complexity to demonstrate how they become efficient as an approximation algorithm. Similar complexity is given in [1]. c) A huge issue in this paper is that VQGS+ is actually not solving the original QOSP problem and thus the 1-1/e or 0.632 bound doesn't apply to VQGS+ and CheetahTraj. It actually solves QOSP problem under a theta constraint similar to [1]. But the argument could be the solved problem is an

	approximation of QOSP problem. The authors should make it clear and provide a bound in the new settings, otherwise it introduces misunderstandings. 2) Presentation: a) Theorem 1 and Lemma 1 should be swapped. Given an algorithm is sub-modular the bound can be proved. Showing 0.632 as the ratio is not accurate, it should be 1-1/e and 0.632 is a rounding number. b) Sec 5, "However, VQGS+ has a long sampling time (e.g., several seconds to tens of seconds) " It worth mentioning the time under what kind of data scale. 3) Lack of novelty: a) Given a prior knowledge of paper [1] some points in the paper is not novel. E.g. the bound 1-1/e, lazy update and visual constraint. Actually a bunch of studies in [3-6] studied the visual constraint. A decent discussion is necessary about why the theta parameter introduced Manhattan distance makes sense. b) The offline optimisation is doing nothing but helps to query and randomly drops trajectories in findRet(). 4) Lack of details: a) In Sec 4.1, a key issue is how to update the uncovered trajectories. While it's not mentioned in the paper. A guess is for each pixel an inverted index is needed for the covering trajectories. The authors should elaborate more. b) There should be more details about the datasets. E.g. what are the data, like taxi trajectories? And also more statics about the dataset. E.g. for a trajectory in average how many pixels are covered? c) In the InvQuad-tree how to decide the height of the tree? d) All the comparison with DTW is under the assumption that the sampling rate is same? 5) Lack of experimental studies: a) [1] should be a comparable baseline as well. b) The choice of theta is not mentioned at all in the paper. It's a key factor to VQGS+ and CheetahTraj. Also it worth studying how theta varies affects the performance. c) Comparing DTW with an offline algorithm is not fair enough. An offline of DTW should be there. d) The offline version of VQGS is also necessary since it preserves the approximate bound.

Review 3

Relevance to SIGIR:	**2**: (Marginally relevant and there's an obviously better venue.)
Quality of Presentation:	**4**: (With all the essential content and understandable by most readers.)
Originality of Work:	**3**: (Somewhat conventional: A number of people could have come up with this if they thought about it for a while.)
Technical Soundness:	**4**: (The authors mostly use appropriate description of technical facts but there are some minor mistakes.)
Impact of Contributions:	**3**: (Interesting but not too influential. The work will be cited, but mainly for comparison or as a source of minor contributions.)
Adequacy of Citations:	**3**: (Just good enough: Review is OK; just about enough past work is covered to make it acceptable.)
Reproducibility of Methods/Findings:	**4**: (Could be reproduced: Researchers could reproduce the methods and results with some effort.)
Overall evaluation:	**-1**: (Weak reject (the paper is near the borderline, but on balance the problems outweigh the strengths, so it should not be accepted)) This paper presents an approach for the visualization of large scale trajectories in an efficient way. The main idea of the paper is to define a visual quality function and two algorithms that solve the sampling problem by exploiting a quad-tree inverted index. Experiments are run on three publicly available dataset and compared with some trivial baselines (random and DTW). The paper is quite well written and the proposed solution has a marginal novelty. The baseline against which it compares to, are quite trivial (e.g random) and it does not compare with other trajectories visualization methods in the literature. Also, it faces a marginal problem for a SIGIR audience and could be more suitable for a Geographical and spatial focussed venue.

Metareview

Metareview for paper 321

Title:	CheetahTraj: Quality and Efficiency in Large-scale Trajectory Data Visual Exploration
Authors:	Qiaomu Shen, Chaozu Zhang, Xiao Yan, Chuan Yang, Dan Zeng, Wei Zeng and Bo Tang
Text:	The paper proposes an approach for visualizing trajectory trace data. Both algorithms and an overall system are presented and evaluated. While reviewers note that the topic is interesting, there are a number of concerns that make the work unsuitable in its current state. I do no think these are easy to remedy in this revision. A central concern for two reviewers is the fit to SIGIR. While I usually have a fairly expansive view of papers that might fit and be of interest to the community, this paper is a bit further a field. One of my tests in situations like this is to see if the work cites either related work at the conference or, at the very least, papers/systems that might benefit from the materials presented. Neither is readily apparent here. It may be possible to make this argument, but it's difficult to imagine. Despite the fit issue, reviewers engaged substantially with the work and provided many additional points of critique (R1 and R2, in particular). I encourage the authors to engage with these as they revise this manuscript. These would also need to be addressed (better comparisons, potentially modified proofs, etc.) and these would be outside the scope of this revision cycle.