SPATIAL DATA ENGINEERING ELECTRONIC REFEREE REPORT

Name of Referee: KYLE SCHUTT
Student ID: 9043-12335
Title: Analysis of Multi-Dimensional Space-Filling Curves

I. SUMMARY AND RECOMMENDATION (TO BE WITHHELD FROM AUTHOR)

Summary of Evaluation
- [] Excellent
- [] Good
- [X] Fair
- [] Poor

Recommendation
- [] Accept without changes
- [X] Accept if certain minor changes are made (see Section IV)
- [] Author should prepare a major revision (see Section IV) for another round of review
- [] Reject

If the paper is accepted, it should be published in SPATIAL DATA ENGINEERING as
- [X] a paper
- [] a short notice

If the paper is rejected, the author(s) should
- [X] Prepare a major revision and resubmit it as a new paper
- [] Submit it to another journal or conference
- [] Regard it as not publishable

II. COMMENTS TO BE WITHHELD FROM AUTHOR – SEE IV

III. OVERVIEW

A) Reader Interest

1. Is the paper of current interest to a reasonable segment of SPATIAL DATA ENGINEERING readership?
- [X] Yes
- [] Perhaps
- [] No

2. Relative to the current level of reader interest in the paper, how is this interest likely to change during the next five years?
- [] Growing interest
- [X] Relatively little change
- [] Diminishing interest

3. Within its particular field of specialization, is the topic of the paper considered important?
- [X] Yes, definitely
- [] Moderately so
- [] Not really

B) Content

1. Is the paper technically sound?
- [] Yes

[X] Appears to be, but didn't check completely
[] Only partially
[] No

2. How would you describe the technical depth of the paper?
[] Expert level
[X] Appropriate for someone working in the field
[] Suitable for the non-specialist
[] Superficial

3. Does the paper make a tangible contribution to the state-of-the-art in it's field?
[] Yes, definitely
[X] To a limited extent
[] No

4. Is the bibliography adequate?
[X] Yes
[] Yes, after certain additions and/or deletions are made (see Section IV)
[] No

5. To what extent is material in the paper likely to be used by other researchers and practitioners?
[] Large
[X] Average
[] Small

C) Presentation

1. Is the abstract an appropriate digest of the work presented?
[X] Yes
[] No

2. Does the introduction clearly state the background and motivation in terms understandable to the non-specialist?
[X] Yes
[] Probably
[] No

3. How would you rate the overall organization of the paper?
[X] Satisfactory
[] Could be improved
[] Poor

4. Relative to its technical content, is the length of the paper appropriate?
[X] Yes
[] No, it should be lengthened
[] No, it should be shortened

5. Is the English satisfactory?
[X] Yes
[] No

6. How readable is the paper for a computer scientist or engineer who is not a specialist in this particular field?
[X] Readable with ordinary effort
[] Paper is self-contained, but a considerable effort is required

[] If the definitions of certain concepts, terms, and symbols were included (noted by "define" in the margins), readability would be improved
[] Less than half the paper is readable
[] Unreadable

7. Disregarding technical content, how would you rate the quality of this presentation?
[] Excellent
[X] Good
[] Fair
[] Poor

IV. DETAILED COMMENTS (TO BE RETURNED TO AUTHORS)

The major contribution of this paper is to provide a simple approach to categorizing specific space-filling curves (SFCs) in such a way that will help scientists understand the architecture of the resulting 1-dimensional space from some D-dimensional space. The authors propose to achieve this by creating a description vector based on several specific (and descriptive) terms: Jump, Contiguity, Forward, Reverse, and Still. While the first four of these terms are self-explanatory based on the movement from one point to another in a specific space, the fourth one, Still, is the closure property that encompasses all other moves that do not fall into one of the distance or direction segments. Still is treated as a closure type but the visualization suggested by the authors does not help describe what Still actually means.

In Figure 3, the authors show the relationships between differing segment types. In the 2-dimensional example, they do not state what type of curve is being used to fill the space. The impression given by the authors is that the relationships are static depending on the curve applied. That being said it is hard to understand where a segment does not follow into some segment type that is not Still. For example, in Figure 3, the cell immediately below the selected cell should be labeled a Jump in a Scan SFC or Sweep SFC because the distance between the two cells is greater than 1. Additionally, it would be labeled as a Forward in both SFCs as well. In either case, the idea that it is a Still makes no sense. The authors could easily expand on showing the relationship based on the type of SFC employed by the reader.

The second issue is that of the time complexity suggested by the authors to calculate the description vector. The authors stated that the time complexity is a uniform $O(N^D)$ for all calculations. This is true for four of the case studies provided by the authors, but it is unconvincing for the fifth case study because it now includes an additional nested summation per dimension based on the size of the grid which means the complexity is closer to $O((N^2)^D)$ for that particular case.

After the initial discussion about the formulations and the layout of the segment types, the author analyzed the scalability, fairness, and intentional bias of each of the segment types against different dimensions and SFCs. While the experiments seemed thorough enough to retrieve potentially useful data, the analysis of this data falls very short of the proposed contribution in the abstract of the paper. The reader is left under the impression that they would be provided with a method for understanding the performance and architecture of a specific SFC, and its implications for implementation. They touched briefly on how certain segment types lend towards more efficient disk operations, but the authors did not bring this up for discussion during their analysis. This occurred for all three analysis sections and the reader is left with fancy graphs showing that Hilbert's SFC is probably the best choice for every problem (even though this is not remotely possible).

Additionally, the analysis of intentional bias may be inaccurate due to the fact that they state that Hilbert's SFC does not change from one dimension to another (i.e., is constant regardless of the dimensionality). However, they state this is due to the fact that the standard deviation (fairness) is low. The problem with this statement is that while the standard deviation is low, it is, by no means, constant. The authors even point out that as dimensionality increases the standard deviation of Hilbert's SFC will, eventually, be worse than other SFCs.

With their analysis complete, the authors come to an abrupt halt in the paper with a short, four sentence paragraph which leaves the reader with no idea of the nature of the contribution. Looking back through the paper, the reader cannot find an actual solution to the description vector describe in each of the case studies. Interestingly, the final description vector is equivalent in every case study even though the segment type functions are vastly different. The reader would be more fulfilled if the authors showed an example calculation of a 2-dimensional Hilbert's SFC on a grid of size 16. This would provide an exact idea about what the description vector looks like and is trying to describe.

Finally, the authors discussed nothing in the way of future work. As a computer scientist or spatial engineer, the reader is concerned about the potential computational inefficiencies outlined by the authors. Specifically, a $O(N^D)$ is computationally expensive. The authors should at least discuss the scalability problem of the equations and functions proposed in the paper as part of a future work section.