Quantify the Effect of Histogram Intersection in Spatio-Temporal Data Sampling

Abstract

- Computational speed is limited by I/O, causing a bottleneck within the system.
- Spatio-temporal sampling reduces the overall data size but is limited by various user inputs, such as histogram threshold.
- We explore the impact of the histogram threshold on Spatio-Temporal sampling over various data sets.
- Results indicate an increase of 100 - 130% in sampling bandwidth with only a 5% decrease in PSNR for the ExaAM dataset.

Spatio-Temporal Sampling

Testing begins with setting the 100% histogram intersection as the base case. Then compare the base case results with 0%, 50%, and 90% histogram intersection to determine the succeeding testing percentage until the percent of blocks reused becomes static.

Datasets

- Asteroid Impact dataset at selected timesteps
- ExaAM Dataset across few selected timesteps
- Hurricane Dataset

Results

The dotted line indicates the histogram threshold at which the ExaAM dataset has improved the sampling bandwidth by 100%, with only a decrease of 5% in PSNR compared to the base case of 100% histogram intersection.

For the figure above, it approaches the maximum blocks reused at the 70% histogram threshold. That is also the most optimum histogram threshold for this dataset. It shows an improvement of 40% in sampling bandwidth with only a 5% decrease in PSNR compared to the base case of 100% histogram intersection.

Conclusion

- Results indicate that the individual dataset limits the improvement of PSNR or sampling bandwidth through the histogram threshold.
- Dataset that falls under the fast-moving categories shows an improvement in sampling bandwidth when histogram intersection decrease. For example, the ExaAM dataset shows an increase of 100-130% for histogram intersections lower than 60%.
- Our methods show that once we approach the static blocks reused region, that is also the region of most optimum histogram intersection.

Acknowledgement

Clemson University is acknowledged for generous allotment of compute time on the Palmetto cluster. This material is based upon work supported by the National Science Foundation under Grant No. SHF-1910197 and SHF-1943114.


Clemson Palmetto Cluster
1 NVIDIA V100 GPU
1 Intel Xeon 6148G CPU

Slow moving data set

Fast moving data set

A fast-moving dataset, such as the ExaAM dataset moves throughout the time step causing the surroundings to change rapidly. These rapid changes in the surroundings result in the improvement in sampling bandwidth as the histogram threshold decreases.

A slow-moving dataset such as the Impact dataset moves throughout the time step without causing many changes in the surroundings. These lack of changes in the surroundings result in no improvement in sampling bandwidth or PSNR as the histogram threshold decreases.

Contact

Changfeng Zou Dr. Jon C. Calhoun (Advisor)