Structured Abstract

Peiqiang Li - 200987503

Context: High quality terapixel visualization rendering is possible through cloud supercomputers. In order to produce a terapixel 3D city visualization supporting daily updates requires very high computational power. Therefore, the performance evaluation of cloud supercomputer architecture becomes particularly important.

Objective: The purpose of this project is to apply exploratory data analysis to study the relationship between the system GPU metrics and rendering performance under the application checkpoint. In addition, the results of data analysis will be used to improve the rendering performance of the cloud computing system.

Method: The project adopts **CRISP-DM** data exploratory analysis process, and **ProjectTemplate** provides a perfect data science development template. `dplyr` of the `tidyverse` package as a tool for efficient data cleaning and processing. `ggplot` is used to draw beautiful graph to show the best results of analysis.

Results: The task scheduling process of GPU takes up a lot of task execution time. The most distinguishing indicator of gpu performance is the memory usage in the running state of the gpu. Different tiles of images have different computational requirements. By establishing the model, the performance of GPU can be classified with GPU serial number to a certain extent.

Novelty: The method of exploratory data analysis has been adopted to analyze the relationship between rendering performance and GPU metrics. In the analysis process, I first compare the overall situation, and then use a specific GPU as an example for comparison. Moreover, machine learning regression models are adopted to distinguish between different performance GPU types.
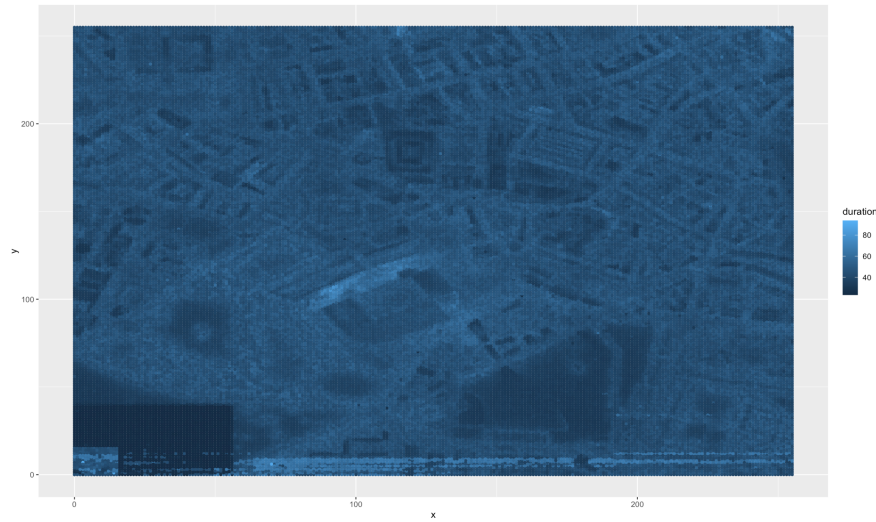
Key Images

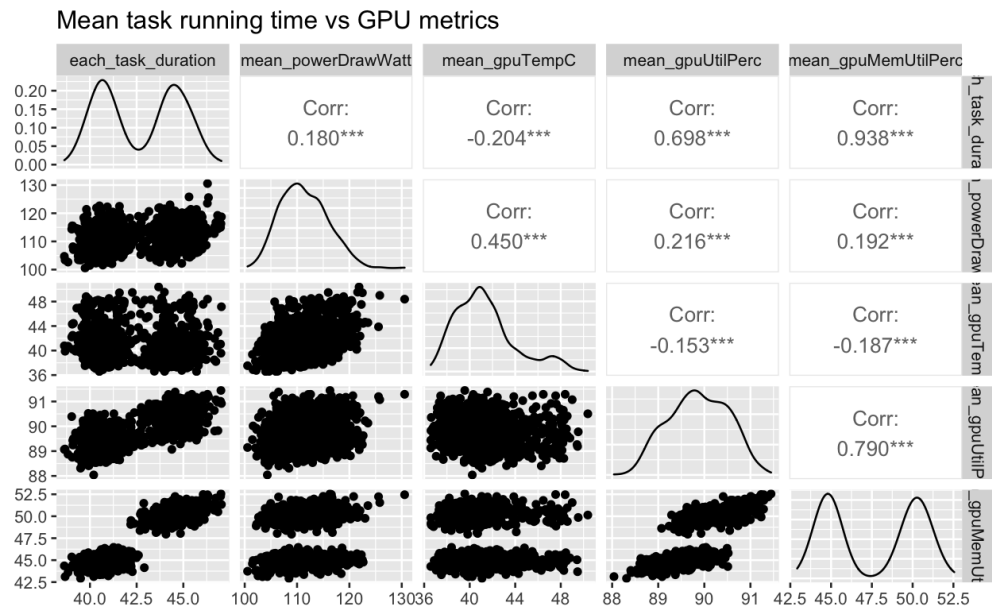Figure 1: Rendering time required for each tile corresponds to the coordinates of the pixel

Figure 2: Mean task running time vs GPU metrics