Dimensionality, Scale and Cognition in Large Dataset Visualisation

Frederick Leighton

September 2016

Dissertation submitted in partial fulfilment for the degree of
Master of Science in Big Data

Computing Science and Mathematics
University of Stirling

Abstract

Problem

Large datasets require non-traditional methods to process, store and analyse them effectively. The same is true when these datasets are visualised, as common plots and graphs cannot fully represent them due to their large scale and high dimensionality. An alternate approach is required to solve the problems that come with size and dimensionality, such as processing efficiency, accuracy of representation, memory requirements and coherent user interaction. These are issues addressed in this project in order to create an informative visual representation of large datasets.

Objectives

The primary objective of this project is to use the Unity3D game engine to create a prototype program which renders large scale datasets in 3-dimensional space. The problems described above must first be defined by performing a comprehensive discussion of relevant issues in current publications and software. The results of the review can then create distinct requirements for the prototype. A successful implementation of the prototype will provide a method by which to gain an informative overview of large datasets.

Methodology

The scientific state-of-the-art in visualisation was reviewed, while ensuring the evaluated work was relevant to 3D, high-dimensional and large scale visualisation. Then an exploration of current software with 3D rendering functionality was completed to identify existing capabilities and shortcomings. Following this, to ensure the resulting prototype is well optimised, the 3D rendering capabilities of Unity3D was tested. The results of the review, discussion and testing then informed a rational design and implementation of the prototype big data visualiser.

Achievements

The main objective of creating a system which renders large datasets in 3D space has been completed. The review of the state of the art and relevant software highlighted issues and considerations when visualising data on this scale. Then, the performance tests showed the performance capabilities of Unity3D and an optimal technique for representing data points in 3D space. The resulting visualiser addresses identified issues and demonstrates its effectiveness in rendering 3D plots. Additionally, an extension of the program was completed to exhibit its extensibility to domain specific datasets, while validating earlier design choices. The prototype successfully fulfils the purpose of this project and provides several solutions to common problems when visualising big datasets in 3D space.