SPACE OPTIMISATION IN MULTI-DIMENSIONAL DATA VISUALISATION

by

Wen Bo Wang

Supervisor: A/Prof. Mao Lin Huang

Co-Supervisor: Dr. Quang Vinh Nguyen

Faculty of Engineering and Information Technology

University of Technology, Sydney

18 April 2016

A thesis submitted in fulfilment for

the degree of Doctor of Philosophy

in

School of Software
And
The Global Big Data Technologies Centre (GBDTC)

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

UNIVERSITY OF TECHNOLOGY SYDNEY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

SIGNATURE OF STUDENT

_____

ACKNOWLEDGEMENTS

How time flies. It has been nearly four years since I first set foot on the University of Technology campus in Sydney, a place that I had been dreaming for years to go. In the precious semesters that followed, I acquired a great deal of knowledge. Indeed, I came to view myself as a tiny boat sailing into the light, away from storms and loaded to full capacity with information, skills, achievements, expectations and dreams. At the tranquil and solemn campus, there is a profound academic atmosphere which encourages rigorous study aspirations. My experience at the University has left me with many happy memories and a strong sense of gratitude in my heart.

Firstly, I would like to gratefully acknowledge A/Prof. Mao Lin Huang who is an expert in the field of information visualisation, and has earned a great reputation at home and abroad for his numerous contributions to scientific research and practical achievements.. This thesis directly grew out of a series of valuable discussions that I had with him.

Secondly, I would like to thank Dr. Quang Vinh Nguyen for guiding me in the various steps towards completing the thesis including research and writing. What deeply impressed me is not just the toil he underwent in numerous readings and modifications, but the wise and noble academic vision he possessed, which will benefit me a lot throughout my life. He brought me closer to the reality I had initially perceived, and eventually enabled me to grasp its rich complexity.

Furthermore, I owe sincere thanks to the members of the Visualisation Team and fellow researchers and the professors of The ***Global Big Data Technologies Centre (GBDTC)****. Their scholarly attainments, instructions, sense of humor, and patient help benefited me in so many ways.*

This research also benefited tremendously from the assistance of many other researchers and staff within the University of Technology . In addition, I would also like thank the participants in the usability study for their cooperation and valuable feedback. Their assistance was as a key factor in providing the necessary

guarantees on the accuracy, objectivity, and precision of the datum I collected.

Last but not least, I owe my deepest gratitude to my brothers and sisters. They have given me so much support and their prayers have enabled me to go through the whole process with joy and peace.

Table of Contents

Figure List

Table List

Equation List

Abstract

Multi-dimensional Data Visualisation (MDV) is the technique to generate visual presentations of datasets with more than three features (or attributes). These graphic representations of data and associated data features can facilitate human comprehension, extraction of implicit patterns and discovery of the relationships among numerous data items for visual data analysis.

Although many optimisation methods have been proposed in the past to improve the visual data processing, not many have been applied to MDV. In particular, little research work has been done in the field of display space optimisation. This thesis focuses on the optimisation of two popular Multi-dimensional visualisation methods: 1) scatterplot matrix and 2) parallel coordinate plots, visualisation by using unique approaches to achieve display space optimisation and interaction.

The first contribution of the thesis is proposing a new visualisation approach named the *Spaced Optimised Scatterplot Matrices* that achieves the maximization of the display space utilisation through position transferring. Breaking through the limitation of discovering the pairwise variable relationships, the new method is able to explore the influences of a single variable towards others. In addition, our algorithms improve the efficiency of interactive Multi-dimensional data visualisation significantly, through the reduction of the computational cost.

The second contribution of the thesis is to improve the parallel coordinate plots and apply it to the computer forensic investigation. As we are living in a big data era, it is much harder for the researchers to provide accurate evidence for victims within a certain time frame. Our research shows that visualisation techniques can improve the working efficiency of investigations in certain cases.

To conclude, we propose a concept of a space optimised Scatterplot Matrix(SPM) visualisation technique considering the shortcomings of the exsited SPM and parallel coordinates in Multi-dimensional visualisation research area. In the meantime, to demonstrate the necessity of our research methodologies, we apply them into computer forensics, which is an area needed analyzing abilities with higher accuracies and efficiencies. By the tests on using Parallel Coordinates and DOITrees, the forensic specialists can easily discover the necesary information in different cases. In the future, we plan to improve our space optimised scatterplot matrix technique from technological optimisation and the broaden application aspects. For example, dealing with the visualisation coolusion problem, non-trivial computation time issue, etc; We will also do the investigations to enlarge the development and availability of Multi-dimensional visualisation techniques.

