

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

Table of Contents	v
Figure List	ix
Table List	xv
Equation List	xvi
Abstract	xvii
Chapter 1 Introduction	1
1.1 Multi-dimensional Data Visualisation (MDV).....	3
1.1.1 Basis of Data Properties	7
1.1.2 Multiple Views.....	14
1.1.3 Scatterplot matrix	16
1.1.4 Parallel Coordinate Plots	23
1.1.5 Other Visualisation Techniques.....	28
1.2 Visualisation Applications of MDV	32
1.2.1 Multi-dimensional Visualisation in Social Science	33
1.2.2 Multi-dimensional Visualisation in Geography.....	34
1.2.3 Multi-dimensional Visualisation in Aerospace.....	35
1.2.4 Multi-dimensional Visualisation in Medicine	36
1.3 Visualisation Optimisation Methodologies in Visualisation.....	38
1.3.1 Modelling Optimisation	39
1.3.2 Geometrical Optimisation	41
1.3.3 Aesthetic Optimisation.....	44
1.3.4 Functional Optimisation.....	46
1.3.5 Applicability Optimisation	47

1.4	Research Challenges	48
1.5	Research Objectives	50
1.6	Contributions	51
1.7	Thesis Organisation	53
Chapter 2 Visualisation Pipelines		56
2.1	Defining Terminology	57
2.2	Data Visualisation Pipelines	60
2.2.1	Data Formatting	61
2.2.2	Data Preprocessing	63
2.2.3	Visual Mapping	66
2.2.4	Human Perception	69
2.3	Visualisation Principles	70
2.4	Summary	70
Chapter 3 Optimisation of Scatterplot Matrix		72
3.1	Framework	73
3.1.1	Idea Evolution	74
3.1.2	Space Optimisation Process	76
3.2	Algorithms	79
3.2.1	Technical Specification	80
3.2.2	Implementation Algorithms	88
3.3	Interaction Mechanism	93
3.3.1	Introduction	94
3.3.2	Interaction Method	95
3.4	Evaluation	99
3.4.1	Performance Evaluation	100
3.4.2	User Studies	102
3.4.3	Supplementary Views	104

3.5	Summary	106
Chapter 4	Interactions with Scatterplot Matrix Visualisation	107
4.1	Dimensionality Reduction	111
4.1.1	Rough Set Theory	114
4.1.2	Variable Precision Rough Set	116
4.1.3	Feature Ranking	119
4.1.4	K-Means for Data Discretisation	120
4.2	Interactive Exploration	121
4.2.1	Point to Region Interaction	122
4.2.2	Linear Approximation of Decision Trend	124
4.2.3	Augmenting Class Coverage	127
4.3	Evaluation.....	129
4.3.1	Case Studies	130
4.3.2	Usability Study	138
4.4	Summary	141
Chapter 5	MDV in Forensic Visualisation	142
5.1	What is Forensic investigation?	144
5.1.1	Digital Evidence	145
5.1.2	Standards and Protocols	146
5.2	MDV Improves Forensic Investigation Models	147
5.2.1	Forensic Investigation Pipeline	148
5.2.2	Visualisation based Forensics' Model	149
5.2.3	A Case Study	152
5.2.4	Summary	155
5.3	MDV Assists Visualisation Hard Disk Drives' (HDDs') Investigation ..	156

5.3.1	Hard Disk Drive	157
5.3.2	Parallel Coordinates' on HDDs	161
5.4	MDV Works in Criminal Relationship Detection	165
5.4.1	Related Works	166
5.4.2	Self-Organizing Map	168
5.4.3	Our Approach.....	174
5.4.4	Experiment and results	179
5.4.5	Conclusion.....	185
5.5	MDV Helps in Crime Analysis	186
5.5.1	Tree Visualisation	187
5.5.2	DOITree Visualisation on Enron Email Dataset	189
Chapter 6 Conclusion and Future Work.....		192
6.1	Reflections on thesis questions	193
6.2	Answers to thesis questions	195
6.3	Future Work	197
6.3.1	Technological Optimisation.....	198
6.3.2	Systematic Scatterplot matrix Evaluation Guidance.....	199
6.3.3	Cooperate with Other Domains	200
Publication List		201
Reference		203

Figure List

FIGURE 1-1 THE CLASSIFICATION OF MULTI-DIMENSIONAL VISUALISATION 1

FIGURE 1-2 DATA VISUALISATION IN SOCIAL MEDIA - VISUALIZING ONLINE CONVERSATIONS; LEFT: A GROUP WITH A SINGLE DOMINANT MEMBER. RIGHT: A GROUP WITH MANY MEMBERS AT DIFFERENT LEVELS OF PARTICIPATION (DONATH 2002). 4

FIGURE 1-3 DATA VISUALISATION IN BIOCHEMISTRY: VISUALISATION OF STRUCTURAL DYNAMICS OF THE RIBOSOME BY DIFFERENT APPROACHES; A) PRESENT CRYO-EM MAP COLOURED ACCORDING TO LOCAL RESOLUTION AS DETERMINED BY RESMAP 2; B) PRESENT CRYO-EM MAP COLOURED ACCORDING TO THE B FACTORS OBTAINED FROM THE PSEUDO-CRYSTALLOGRAPHIC ATOMIC MODEL REFINEMENT (FISCHER ET AL. 2015)..... 4

FIGURE 1-4 DATA VISUALISATION IN NETWORK SECURITY: EIGHT VERTICAL AXES REPRESENTS A 2.5 CLASS B IP RANGE(SHIRAVI, SHIRAVI & GHORBANI 2012)..... 5

FIGURE 1-5 DATA VISUALISATION IN SCIENTIFIC PROJECT TRIDENT: IT TRANSFORMS RAW DATA FROM SENSORS INTO VISUALISATIONS AND DATA PRODUCTS.(BARGA ET AL. 2008)..... 5

FIGURE 1-6 DISPLAY THE NEWSPAPER ENDORSEMENTS IN THE US PRESIDENTIAL ELECTION IN 2004. SOURCE FROM: [HTTPS://FLOWINGDATA.COM/2008/10/29/MAP-SHOWS-NEWSPAPER-ENDORSEMENTS-IN-US-PRESIDENTIAL-ELECTION](https://flowingdata.com/2008/10/29/map-shows-newspaper-endorsements-in-us-presidential-election/) / 8

FIGURE 1-7 ORDINAL DATA VISUALISATION: A GRAPH DESCRIPTION OF THE USERS’ PREFERENCE. SOURCE FROM: 8

FIGURE 1-8 QUANTITATIVE DATA VISUALISATION: A) VISUALLY MONTHLY HEALTH DATA BY MEANS OF PENCIL ICONS ON A MAP (TOMINSKI, SCHULZE-WOLLGAST & SCHUMANN 2005); B) AN ANIMATED FLOW VISUALISATION CREATED FROM STREAMLINE IMAGES (VAN WIJK 2002)..... 9

FIGURE 1-9 A VISUALISATION SAMPLE FOR LINEAR STRUCTURE: IT SHOWS GLOBAL AVERAGE TEMPERATURE FROM FIVE DATASETS SINCE THE START OF THE SATELLITE TEMPERATURE DATA ERA IN 1979 THROUGH 2009. SOURCE FROM: [HTTP://APPINSYS.COM/GLOBALWARMING/LINEARTRENDS.HTM](http://appinsys.com/globalwarming/lineartrends.htm)..... 10

FIGURE 1-10 DATA TREE STRUCTURE. 11

FIGURE 1-11 AN EXAMPLE OF A SPACE OPTIMISED TREE VISUALISATION, THE DATASET CONTAINS OF 11

FIGURE 1-12 AN EXAMPLE FOR NETWORK STRUCTURE. DETECTING CRIMINAL ORGANIZATIONS THROUGH VISUALISATION

(ROBERTSON, MACKINLAY & CARD 1991).	12
FIGURE 1-13 MULTIPLE VIEWS OF THE 6D SAMPLE DATASET SHOWN IN TABLE 1	15
FIGURE 1-14 A SCATTERPLOT MATRIX VISUALISATION SAMPLE.....	16
FIGURE 1-15 SCATTERPLOT MATRIX VARIETIES – INFORMATION ENRICHMENT: THE UPPER-LEFT IS USING HISTOGRAM TO REPLACE THE PLOTS IN THE DIAGONAL; THE UPRIGHT IS USING DATA TREND TO REPLACE THE PLOTS IN THE DIAGONAL, AND USES DIFFERENT VISUAL MARKS TO EXPLAIN THE WHOLE DATASET; THE DOWN-LEFT IS A VISUALISATION METHOD CALLED HYPERSLICE (VAN WIJK & VAN LIERE 1993), WHICH IS A MATRIX OF PLOTS WHERE “SLICES” OF MULTIVARIATE FUNCTION ARE DISPLAYED AT A CERTAIN FOCAL POINT OF INTEREST; THE DOWN-RIGHT USES VARIABLE NAME TO REPLACE THE DIAGONAL PLOTS, AND USES COLOUR TO DISPLAY THE DATASET’S CATEGORY.	17
FIGURE 1-16 BRUSHED SCATTERPLOTS.....	18
FIGURE 1-17 GENERALIZED SCATTERPLOT MATRIX	19
FIGURE 1-18 POLYGON SCATTERPLOTS.....	19
FIGURE 1-19 DIMENSION REORDERING SCATTERPLOT MATRIX	20
FIGURE 1-20 3D SCATTERPLOT MATRIX SHOWING THE 8D “OLIVE OIL” DATASET. (SANFTMANN & WEISKOPF 2012)	21
FIGURE 1-21 GEOMETRY THEORY OF PARALLEL COORDINATE IN 2D	23
FIGURE 1-22 PARALLEL COORDINATES PLOT FOR THE CAR DATASETS	25
FIGURE 1-23 CAR DATASET VISUALISED BY ARC-BASED PARALLEL COORDINATES GEOMETRY (HUANG, LU & ZHANG 2015).....	26
FIGURE 1-24 CARS DATASET VISUALISATION IN PARALLEL COORDINATES WITH A NEW AXES RE-ORDERING METHOD (LU, HUANG & HUANG 2012).....	26
FIGURE 1-25 FORBES 94, A DATASET WITH 5 VARIABLES VISUALISED IN PARALLEL COORDINATE VISUALISATION: AFTER CLUTTER REDUCTION (LU ET AL. 2010).	27
FIGURE 1-26 EXAMPLE OF GRAPHICAL LOGICAL DIAGRAMS(TAYLOR ET AL. 2006).....	29
FIGURE 1-27 THE THEORY OF THE PIXEL-ORIENTED VISUALISATION TECHNIQUE(KEIM 2000).....	30
FIGURE 1-28 A VISUALISATION SAMPLE OF CIRCLE SEGMENTS TECHNIQUE (ANKERST 2001).	30
FIGURE 1-29 CHERNOFF FACES VISUALISATION FOR LAWYERS’ RATINGS OF TWELVE JUDGES.....	31
FIGURE 1-30 A VISUALISATION APPLICATION IN SOCIAL SCIENCE (GORBAN & ZINOVYEV 2010)	33
FIGURE 1-31 AN EXAMPLE OF MULTI-DIMENSIONAL VISUALISATION IN GEOGRAPHY – 3D TERRAIN PROFILE	34

FIGURE 1-32 AN EXAMPLE OF MULTI-DIMENSIONAL DATA VISUALISATION IN AEROSPACE (HURTER, TISSOIRES & CONVERSY 2010)..... 35

FIGURE 1-33 THE EXAMPLES OF MULTI-DIMENSIONAL DATA VISUALISATION IN GENE ANALYSIS (PAVER); (A) LEVEL REPRESENTATION AND (B) ANALYSIS OF THE GLOBAL MRNA OF A BACILLUS SUBTILIS 168 GLUCOSE STARVATION EXPERIMENTANALYSE (OTTO ET AL): SOURCE FROM: [HTTP://WWW.DECODON.COM/PAVER-BENEFITS.HTML](http://www.decodon.com/paver-benefits.html)..... 37

FIGURE 1-34 THE RATIONALE OF THE ARC COORDINATES PLANE..... 42

FIGURE 1-35 (UP) ORIGINAL PLOT (HAUCK ET AL.); (DOWN) AFTER CLUTTER REDUCTION. DATA RESULTS USED WITH KIND PERMISSION OF THE AUTHOR LIANG FU LV. 43

FIGURE 1-36 CIRCULAR TREEMAPS EXAMPLE, THE RED CIRCLES REPRESENT NEWER FILES WHILE SOFT YELLOW CIRCLES REPRESENT OLDER FILES. SOURCE FROM: [HTTP://LIP.SOURCEFORGE.NET/CTREEMAP.HTML](http://lilp.sourceforge.net/ctreemap.html) 44

FIGURE 1-37 AN EXAMPLE OF COMPUTER SCREEN THAT ACHIEVES THE MAXIMIZATION (100%) OF SPACE UTILISATION AND THE MINIMIZATION (0%) OF THE OVERLAP AMONG TWO SESSION DISPLAYS BY USING THE NEW TANGRAM TREEMAPS (LIANG.J 2015) 45

FIGURE 1-38 DATA CLUSTERING IN PARALLEL COORDINATE. CASE STUDY WITH WAGES DATASET OBTAINED FROM [HTTP://WWW.NBER.ORG/CPS/](http://www.nber.org/cps/). 46

FIGURE 1-39 CHERNOFF FACES DISPLAY LIFE IN LOS ANGELES. SOURCE FROM: 47

FIGURE 1-40 THESIS STRUCTURE DESCRIPTION..... 53

FIGURE 2-1 VISUALISATION PROCESS..... 60

FIGURE 2-2 CROSSFILTER VISUALISATION TOOL DISPLAYS THE AIRLINE ON-TIME PERFORMANCE. SOURCE FROM: [HTTP://SQUARE.GITHUB.IO/CROSSFILTER/](http://square.github.io/crossfilter/) 63

FIGURE 2-3 AN EXAMPLE OF A CLUSTERING HEAT MAP. THE ROWS ARE THE HIERARCHICALLY CLUSTERED GENES, WHILE THE COLUMNS ARE THE TISSUES WITH DENDROGRAMS. THE RED IN THE HEAT MAP PRESENTS UPREGULATION WHILE BLUE PRESENTS DOWNREGULATION. CITED FROM: [HTTP://ALTANALYSE.BLOGSPOT.COM.AU/2012/06/HIERARCHICAL-CLUSTERING-HEATMAPS-IN.HTML](http://altanalyse.blogspot.com.au/2012/06/hierarchical-clustering-heatmaps-in.html) 64

FIGURE 2-4 PCA OF A MULTIVARIATE GAUSSIAN DISTRIBUTION CENTERED AT (1,3) WITH A STANDARD DEVIATION OF 3 IN ROUGHLY THE (0.878, 0.478) DIRECTION AND OF 1 IN THE ORTHOGONAL DIRECTION. THE VECTORS SHOWN ARE THE EIGENVECTORS OF THE COVARIANCE MATRIX SCALED BY THE SQUARE ROOT OF THE CORRESPONDING EIGENVALUE, AND

SHIFTED SO THEIR TAILS ARE AT THE MEAN, SOURCE FROM:

[HTTPS://EN.WIKIPEDIA.ORG/WIKI/PRINCIPAL_COMPONENT_ANALYSIS](https://en.wikipedia.org/wiki/Principal_Component_Analysis)..... 65

FIGURE 3-1 PLOT DRAWING-POSITION OF EACH VERTEX..... 77

FIGURE 3-2 WINE DATASET WITH FOUR ATTRIBUTES IS SHOWN IN DIFFERENT SIZE OF DISPLAY SCREEN (UP-LEFT) IS THE FULL SCREEN DISPLAY; (UP-RIGHT) IS THE 50% SCREEN DISPLAY;(DOWN-LEFT)IS THE 25% SCREEN DISPLAY; (DOWN-RIGHT) IS THE 12.5% SCREEN DISPLAY..... 77

FIGURE 3-3 TRADITIONAL SCATTERPLOT MATRIX LAYOUT..... 81

FIGURE 3-4 THE LAYOUT OF A MATRIX WITH PLOTS REDUCTION..... 82

FIGURE 3-5 A EXAMPLE OF A DRAW 84

FIGURE 3-6 A) THE EXAMPLE OF MOVING A DRAW; B) THE PROCESSES OF MOVING A DRAW 86

FIGURE 3-7 A) THE SAMPLE OF ROTATING A DRAW (B) THE PROCESSES OF ROTATING $\pi/4$ OF A DRAW 87

FIGURE 3-8 THE REPOSITION METHOD FOR SCATTERPLOT MATRIX 88

FIGURE 3-9 THE ALGORITHM OF LAYOUT OPTIMISATION APPROACH..... 89

FIGURE 3-10 THE ALGORITHM HIGHLIGHTING THE APPROACH TO REFLECTING THE VARIABLE RELATIONSHIPS..... 90

FIGURE 3-11 (A)(B): THE SAMPLE OF USING COLOUR PROPERTIES TO DISCOVER PAIRWISE VARIABLE RELATIONSHIPS. DATASET DESCRIPTION: THIS DATASET OBTAINED FROM UC IRVINE MACHINE LEARNING REPOSITORY, IT CONTAINS 12 VARIABLES, INCLUDING FIXED ACIDITY, VOLATILE ACIDITY, CITRIC ACID, RESIDUAL SUGAR, CHLORIDES, FREE SULFUR DIOXIDE, TOTAL SULFUR DIOXIDE, DENSITY, PH, SULPHATES, ALCOHOL, QUALITY; WEBSITES: 91

FIGURE 3-12 THE INTERACTION METHOD – USER QUERY 96

FIGURE 3-13 THE USER QUERY RESULT WITH COLOUR PROPERTY 96

FIGURE 3-14 THE USER QUERY RESULT WITH SHAPE PROPERTY..... 97

FIGURE 3-15 (A) (B) (C): EXAMPLES OF SHAPE PROPERTY WORKING ON THE LAYOUT OF THE DATASETS: A) REPRESENTS THE DATA POINTS CONNECTED INTO DIFFERENT SHAPES, AND THIS HELPS TO DISTINGUISH THE OVERALL DISTRIBUTION FOR EACH PLOTS; B) THE GREEN LINE IS THE DIAGONAL LINE OF EACH PLOT, AND FROM THE GAP BETWEEN THE OTHER LINES AND DIAGONAL LINE, IT PROVIDES ANOTHER OVERVIEW OF THE DATASET DISTRIBUTION WITHOUT ANY POINTS DISPLAYED; C) WE USE THIS VIEW TO DETECT THE DENSITY OF DATA POINTS DISTRIBUTION. IT IS MORE CONVENIENT TO FIND THE CENTRE POINT IN A CIRCLE SHAPE COMPARED WITH DRAWING IN A PLOT (RECTANGULAR), BUT AN IMPORTANT FACTOR IS

THAT THE CIRCLE SHOULD BE TANGENTIAL WITH EDGES OF EACH RECTANGULAR (PLOTS).....	98
FIGURE 3-16 (A) THE SAVING NUMBER OF SCATTERPLOTS B): THE SAVING SPACE RATE OF OUR NEW SCATTERPLOT MATRIX.....	101
FIGURE 3-17 THE RESULTS OF USER PREFERENCE	102
FIGURE 3-18 AN OVERVIEW COMPARISON OF PLOT ARRANGEMENTS IN THE NEW LAYOUT AND THE ORIGINAL LAYOUT	104
FIGURE 3-19 DIFFERENTIATION IN COLOUR; EXPERIMENT WITH ABOLONE DATASET	105
FIGURE 3-20 DIFFERENTIATION IN COLOUR AND SHAPE; EXPERIMENT WITH ABOLONE DATASET	105
FIGURE 4-1 A ILLUSTRATION OF SCATTERPLOT MATRIX VISUALISATION	122
FIGURE 4-2: INTERACTION (MOUSE CLICK) BY USING POINT-TO-REGION CONCEPT: THAT IS, A POINT CLICK CAUSES AN ENTIRE CONVEX HULL (A CLASS) TO BE HIGHLIGHTED.....	123
FIGURE 4-3 (A) A CLASSIC SCATTERPLOTS VISUALISATION. (B) ADDING THE DECISION FLOW WHERE PLOTS WERE AUGMENTED WITH RESPECT TO THE DECISION VARIABLE.....	125
FIGURE 4-4 VISUALISATION OF THE ENTIRE WINE DATASET USING A) PARALLEL COORDINATE AND B).....	131
FIGURE 4-5: RESULTS OBTAINED FROM THE CASE STUDY WITH WINE DATA. THE UPPER DIAGONAL	133
FIGURE 4-6: CASE STUDY WITH WINE DATASET OBTAINED FROM [31]. BOXES AT “WITHOUT-TREND” AREA (AREA ABOVE THE DIAGONAL LINE) ARE SCATTERPLOTS OF EACH PAIR OF ATTRIBUTES WHILE BOXES AT “WITH-TREND” AREA (AREA BELOW THE DIAGONAL LINE) REPRESENT THE SAME VALUES AND WITH CHANGING TRENDS.....	134
FIGURE 4-7: CASE STUDY WITH CAR DATASET. WE SELECTED MILEAGE PER GALLON (MPG) AS THE DECISION AND THE DATASET HAS BEEN REDUCED TO 4 ATTRIBUTES NAMEDLY ACCELERATION, DISPLACEMENT, CYLINDERS AND HORSEPOWER.....	136
FIGURE 4-8: CASE STUDY WITH WAGES DATASET. THIS FIGURE SHOWS A BOX AT “WITH-TREND” AREA FOR CORRELATION BETWEEN <i>EXPERIENCES</i> , <i>WAGE</i> (Y-AXIS) AND <i>AGE</i> (X-AXIS).....	137
FIGURE 4-9: ACCURACY OF PARALLEL COORDINATE AND SCATTERPLOT MATRIX VISUALISATIONS CORRESPONDING TO FIVE QUESTIONS (WITH 95% CONFIDENCE INTERVALS).....	139
FIGURE 5-1: A VISUALISATION TECHNIQUE BASED FORENSIC INVESTIGATION MODEL	149
FIGURE 5-2 NEW MODEL FOR HDD VISUALISATION.....	152
FIGURE 5-3: DATA STORAGE CATEGORY.....	157
FIGURE 5-4: THE VISUALISATION OF ORIGINAL METADATA BY PARALLEL COORDINATE	163
FIGURE 5-5 USE PARALLEL COORDINATE TO IDENTIFY THE SUSPICIOUS FILE IN HDD.....	163

FIGURE 5-6: THE THEORY OF THE SELF ORGANIZING MAP (VESANTO & ALHONIEMI 2000)..... 169

FIGURE 5-7: THE DISTANCE BETWEEN TWO NEURONS..... 171

FIGURE 5-8: AN OVERVIEW OF THE WHOLE 16 ITEMS' FEATURE VALUE AND THE CIRCLED AREAS ARE THE SUSPICIOUS DATA. 180

FIGURE 5-9: THE DIFFERENCE BETWEEN ITEM 1, ITEM 4 AND ITEM 16. 181

FIGURE 5-10: THE WEIGHT PLANES OF ALL 16 ITEMS BY SOM 182

FIGURE 5-11: THE CLASSIFICATION HITS AFTER USING SOM OF 16 ITEMS 183

FIGURE 5-12: THE VISUALISATION OF CLASSIFICATION OF ALL FEATURES..... 183

FIGURE 5-13: EXAMPLES OF THE EMAILING ACTIVITIES FOR EIGHT EMPLOYEES IN ENRON. A) CHECK THE ACTIVITIES OF
BAUGHMAN-D. B) USING COLOUR IN DOITREE 190

Table List

TABLE 1-1	A 6D SAMPLE DATASET	14
TABLE 2-1	CANONICAL DATA MODEL - WHITE WINE QUALITY DATASET	61
TABLE 2-2	VISUAL MARKS.....	66
TABLE 2-3	VISUAL PROPERTIES OF MARKS	67
TABLE 5-1	NAME AND TYPE	153
TABLE 5-2	NAME AND REMARK.....	153
TABLE 5-3	TIME ATTRIBUTES.....	159
TABLE 5-4	DISK INVESTIGATOR.....	162

Equation List

EQUATION 1	THE NUMBER OF PLOTS.....	83
EQUATION 2	THE NUMBER OF PLOTS PER ROW	83
EQUATION 3	THE NUMBER OF PLOTS PER COLUMN	83
EQUATION 4	THE RELATIONSHIP BETWEEN THE NUMBER OF SCATTERPLOTS AND THE NUMBER OF VARIABLES	100
EQUATION 5:	THE INDISCERNIBILITY RELATION AMONG OBJECTS	114
EQUATION 6:	THE β POSITION REGION IN VPRS MODEL.....	116
EQUATION 7:	THE QUALITY OF CLASSIFICATION IN VPRS MODEL.....	117
EQUATION 8:	CALCULATE THE RANKING COEFFICIENT FOR EACH CONDITIONAL ATTRIBUTE	119
EQUATION 9:	THE CALCULATION OF THE SLOPE TO MEASURE THE CHANGE	126
EQUATION 10:	THE CALCULATION OF THE INTERCEPT.....	126
EQUATION 11:	INTERPOLATE THE BEST FITTING LINE AT POINT (X_0, Y_0)	126
EQUATION 12:	CALCULATE THE ACCURACY OF A RULE.....	127
EQUATION 13:	CALCULATE THE COVERAGE OF A RULE.....	128
EQUATION 14	CALCULATION OF \hat{X}	169
EQUATION 15	CALCULATION OF \hat{W}_j	169
EQUATION 16	EUCLIDEAN DISTANCE CALCULATION.....	170
EQUATION 17:	MEASURE THE DISTANCE BETWEEN TWO OBJECTS.....	171
EQUATION 18:	CALCULATE THE DISTANCES BETWEEN EVERY TWO OBJECTS.....	172
EQUATION 19:	CALCULATE THE MINIMUM DISTANCE AMONG ALL DISTANCES	172
EQUATION 20:	CERTIFY THAT THE WINNER NEURON IS THE LARGEST SCALAR PRODUCT	172
EQUATION 21:	RECOGNISE THE WEIGHT.....	173

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 existed 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 necessary 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.