Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

VISUAL ANALYTICS: DESIGNING FLEXIBLE FILTERING IN PARALLEL COORDINATE GRAPH

Z. Idrus[1,*], H. Zainuddin[2] and A. D. M. Ja'afar[1]

[1]Faculty of Computer and Mathematical Sciences, UniversitiTeknologi MARA, 40450ShahAlam, Selangor, Malaysia

[2]Faculty of Applied Sciences,UniversitiTeknologi MARA, 40450 Shah Alam, Selangor, Malaysia

Published online: 17 October 2017

ABSTRACT

Data visualization is a technique of creating visual image to help users in speed understanding of big data. The visualization reveals hidden knowledge through data patterns and relationships. The process is known as visual analytics. Nevertheless, complex and huge data created visualization that is congested and hardly reveal the data patterns. Thus, filtering is a technique that gives flexibility to users to control over the data view as such to focus only on interest part and hide others. However, most of the filtering is not studied in a structured manner. Thus, this research has designed a structured process for formulating filtering technique in a parallel coordinate graph since it is a widely used technique for visualizing multivariate data. With the process, flexible visual filtering presentations for parallel coordinate graph have been produced. The finding support a wide range of visual analytics needs in parallel coordinated graph.

Keywords: data visualization; visual analytics; parallel coordinate graph;data filtering.

Author Correspondence, e-mail: zainura@tmsk.uitm.edu.my

doi: http://dx.doi.org/10.4314/jfas.v9i5s.3

1. INTRODUCTION

Data visualization is a term used for visual images that help users to understand complex data [1] through their patterns and relationships [2]. Others define data visualization as the art of conveying information visually instead of presenting them in numerical format [3]. Some of the common form of data visualizations are tables, diagrams, images [4], plots, graphs and charts. The visual image is also essential for complex analyses and the process is defined as visual analytics [5]. It is a science of analytical reasoning and must be built with integrative features[6]. Visuals analytic assists in the process of extracting patterns and relationships that exist within the data[7]. However, in complex and huge data, it is common for patterns to be cluttering that lead to unjustifiable analysis. Thus, filtering is one of the techniques to reduce visual clutter [8-9], data complexity and simplify data patterns, thus create apparent relationships.

Filtering is a process to limit the number of data entity to be displayed as a mean to reduce data congestion. Filtering gives users flexible control over the visualization as such to focus on their interest and hide insignificant items. It is a vital process in visual analytics since it is unusual to visualize and analyze huge data at once [10]. However, most of the filtering is not studied in a structured manner. Thus, this research studies filtering of parallel coordinate graph in a structured approach. The graph is chosen as it is a widely used visualization technology in analyzing multivariate data. The research introduces a systematic process to flexible filtering technique. At the end of the process, a range of flexible visual filtering designs are produced. The finding will support visual analytics needs in parallel coordinated graph.

The remainder of this paper is structured with section 2 highlights the background study of visual analytics in parallel coordinate graphs. Section 3 describes the process of visual analytics design. It also explores parallel coordinate graph as part of the process. Then, an abstraction design of parallel coordinate graph via gold directed approach is undertaken. By adopting abstract scene analysis method, section 4 discusses the seven filtering designs extracted from the graph abstraction. Section 5 detailed out the evaluation of the filtering designs. Finally, section 6 concludes the finding with suggestions for future enhancement.

2. BACKGROUND STUDY

The rationale behind data visualization is to convey information through visual image such as graphical tools. Since data are huge with different levels, data visualization must be equipped with visual analytic features. Visual analytic is also commonly known as visual exploration[11-12]. Some of the common visual analytics techniques are filter, zoom, sort, brush, bind and range[13].Information Seeking Mantra which was introduced by [14] highlights the concept of three steps pathof visual analytics. First step is overview firstwhere data is viewed in a single graph. Visualization technology is useful in this step. Second step is to zoom and filter which is to identify interest patterns and relationships. Finally, detail on demand which is to focus on the interested parts of the data.This is the stage where visual analytic technology becomes vital. Visual exploration activities are performed to view information from different perspectives.

Parallel coordinate graph is one of the visualizationtechnologies. The main strength of the graph is on its ability to reveal the relationship of multivariate data[15-16]in a single graph through its multi-dimensionalaxes. When compared to other visual technology, parallel coordinate graph is well ahead in term of time taken to analyze its data[17]. Parallel coordinate alsowork well with many of visual analytic technology such as filtering, sorting, zooming, slicing, dicing and brushing[18-19].

Beside data, edges of parallel coordinate graph also gain researchers interest. For example, in the study by [11], the edges are divided into three which are within, between and background.The edges can beselected to meet users' needs. For every selection within and between edges are shown while background edges are hidden.

On the other hand, edges in[8] is viewed form visual interaction perspectives. The research introduces a force model which reduces clustering through edges interaction. The force is the main function responsible to reduce the interference between edges by allowing them to curve and adjusts their shape.

In term of methodology, Visualization Pipeline methodology is introduced and filtering is one the steps and it is user-centered [20]. Users decide what variables to be focused,while hide others during exploration.

Similarly, a new methodology to retain the visual exploration abilities of filtering in parallel coordinate graph has been formulated by [9]. The model extends the capability of the current methods by introducing parallel processing. With the new capability, the typical web-based visualization library is now become a client-server model which has the ability to improve plot rendering to 20 times faster

Similarly a model of four functions has been designed for computer forensics investigators. One of the model's functions is filtering. It is applied to sort out unrelated data of crime scene. The aimis to produce speed analysis and discover evidence through interactive visual exploration. [15]. Even though filteringtechnique is widely applied to big data analysis, its various modesof presentations are hardly studied in detail.

3. ABSTRACT DESIGN

We propose a new two structured stepsfor deriving variations in visual analytics presentations. The first step is to understand the graph's behaviors. This approached is adopted form goal directed approach which can be traced backed since 1990 [21]. The graph behaviors such as dimension, concept, attributes type and attributes behaviors are studied. The behaviors and their relationships form an abstraction of the graph. The second step is to identify the activities that can be performed on the abstraction. The process is adopted form abstract scene analysisapproach [22]. With such focused in mind various exploration designscan be derived and expended by domain experts. Thus, the design can be adapted to many applications.

The two steps visual analytics design is implemented to parallel coordinate graph. The first step is to derivethe behavior of parallel coordinate graph.Let G be the parallel coordinated graph. The main entity of the graph areaxes whichare denoted as C_x where

$$C_x = \{1.....C_x\} \tag{1}$$

and it indicates the total numbers of axes in the graph.

During exploration, the axes haveto be in one of two states. They are either selected or not selected to be part of the graph visualization.The state isdenoted as S = {selected, not-selected}. The axes will appear on graph if it is selected. However, at any given time, G has at least one axis selected.

$$G: C_x \rightarrow S \tag{2}$$

Plotted on the axes are dataset and is denoted byD_y where $D = \{1.....D_y\}$. The data can also be in the two modes which is selected or not-selected. Thus,$S = \{selected, not\text{-}selected\}$.At any given time, at least one data is selected.Thus, $G$is a function indicating the state of axes,$C_x$and data, D_y. Thus,

$$G: C_x \, x \, D_y \rightarrow S \tag{3}$$

The position of the axes,Px is not fix and the total number of P is equal to the total number of $C_x$thus $P_x = \{1...P_x\}$where

$$G: C_x \times D_y \rightarrow S \times P \quad (4)$$

To refer to the element of *S* and *P*, *".s"* and *".p"* will be used respectively. *G* is said to be filtered by axis if at least one of its axes $C_x$in *G* is in a *not-selected* mode (regardless ofC_x *position*and *Dy mode*)

$$\exists i \in C_X (G(i).s = not\text{-}selected) \quad (5)$$

Or at least one of the axes in*G* has been assigned to a new position(regardless of C_x *mode* and D_y *mode*)

$$\exists j \in C_X (G(j).p = P_x) \quad (6)$$

Or at least one of the data D_y in *G* is in a *not-selected* mode (regardless of C_x *position* and *mode*)

$$\exists k \in D_y (G(k).s = not\text{-}selected) \quad (7)$$

Let range *R* be a set of *R = {solo, individual, continuous}*

Thus,

$$G: C_x \times D_y \rightarrow S \times P \times R \quad (8)$$

Since this research is interested in the changes of parallel coordinate graph over time, a time variable is added to the function *G* thus it is having the domain of

$$G(t): T \times C \times D \rightarrow S \times P \times R$$

where

$$T = [t_{start}, \infty) \quad (9)$$

wheret_{start} is the starting time for visual analytics process. *G(t)* is the state of the parallel coordinate graph at time *t*.

At any particular time *t*, action will be made to axes. There are three possible actions. First is focus on the axes where they can be in two states which are selected or not selected. Second is when some of the data are selected while other are not selected to be viewed. Finally, change the position of coordinate. The action is denoted by function act:

$$act: T \times C \times D \rightarrow S \cup P \cup R \cup \{no\_change\} \quad (10)$$

no_change to indicate that there is no action has been made to the graph

Let $A(t_x, t_y)$ be the graph of *act* and time *t* is limited to $[t_x, t_y)$ such as

$$A(t_x, t_y) = \{< t, c, d, act(t, c, d) > \mid t_x < t < t_y, c \in C, d \in D\} \quad (11)$$

Thus, flexible filtering design is a temporal function ψ compute all the filtering performs on the graph as such:

$$G(t) = \psi(t, A(t_{start, t})) \quad (12)$$

The function is in fact is an abstract of the parallel coordinate graph supporting it in two vital attributes, which are itsbehaviors that make up the graph and the filtering activities to be performed on its behaviors. Thus, the next section is to identify various filtering design through abstract scene analysis technique [22].

4. VARIATION OF FILTERING DESIGN

From the abstraction of parallel coordinate graph above, the second step is to extract the variation in the filtering to produce flexible filtering designs.

4.1.Scenario 1: Zero Filtering

Parallel coordinate graph display all the axis and dataset.

Abstract Scene Analysis:

All axesC_x and dataset $D_y$are set to *selected* mode. The position P_x is unchanged.

$\forall i \in C_x,\ \forall j \in D_y (G(ti)(i, j) \leftarrow \langle s \rangle)$, where s = selected$\in$ Sfor the duration of *ti*to *tj*where $tstart \leq ti < tj < tend < \infty$ and ti , tj, $tstart$, $tend \in T$ (13)

4.2. Scenario 2: AxisChoice

Axes are active with open selection. However, their positions are static.

Abstract Scene Analysis:

$\exists i \in C_x,\ (G(ti)(i) \leftarrow \langle s, p, r \rangle)$, where s = *selected* $\in S$, $G(i).p \neq p$, $p \in P$, $r \in R$for the duration of *ti*to *tj*where $tstart \leq ti < tj < tend < \infty$ and ti , tj, $tstart$, $tend \in T$ (14)

4.3. Scenario 3: AxisUltimate

Given an axis, it can view and its position can also change.

Abstract Scene Analysis:

$\exists i \in C_x,\ (G(ti)(i) \leftarrow \langle p, r \rangle)$, where s = *not-selected* $\in S$, $G(i).p = p$, $p \in P$, $r \in R$for the duration of *ti*to *tj*where $tstart \leq ti < tj < tend < \infty$ and ti , tj, $tstart$, $tend \in T$ (15)

4.4. Scenario 4: Dataset Ultimate

Only data are selected. There are two types of selection which are range and selection. Range is to select two or more data continuously. While,Solo is to select only one set of data only.

4.5.Scenario 5: Dataset Range

It supportsinteractive selection of dataset.

Abstract Scene Analysis:

$\exists j \in D_y\ (G(ti)(\ j)\ \leftarrow \langle s,\ r \rangle)$, where $s = selected \in S$, $r = \{individual,\ continuous\} \in R$ for the duration of $ti$to $tj$where $tstart \leq ti<tj< tend<\infty$ and $ti\ ,\ tj,\ tstart,\ tend \in T$

(16)

4.6. Scenario 6: Dataset Solo

Only one dataset are allowed to be highlighted, while others are deemed and appear as data background. The data background is important as it creates awareness [23].

Abstract Scene Analysis:

$\exists! j \in D_y\ (G(ti)(\ j)\ \leftarrow \langle s,\ r \rangle)$, where $s = selected \in S$, $r = solo \in R$ for the duration of $ti$to $tj$where $tstart \leq ti<tj< tend<\infty$ and $ti\ ,\ tj,\ tstart,\ tend \in T$ (17)

4.7. Scenario 7: Continuous

The selection of data is nested within the axes. The range of selection is continuous.

Abstract Scene Analysis:

$\exists I \in C_x,\ \exists j \in D_y\ (G(ti)(I,\ j)\ \leftarrow \langle s,\ r \rangle)$, where $s = selected \in S$, $r = continuous \in R$for the duration of $ti$to $tj$where $tstart \leq ti<tj< tend<\infty$ and $ti\ ,\ tj,\ tstart,\ tend \in T$ (18)

4.8. Scenario 8: Split

Unlike Scenario 8, the range selection is not continuous.

Abstract Scene Analysis:

$\exists i \in C_x,\ \exists j \in D_y\ (G(ti)(i,\ j)\ \leftarrow \langle s,\ r \rangle)$, where $s = selected \in S$, $r = individual \in R$for the duration of $ti$to $tj$where $tstart \leq ti<tj< tend<\infty$ and $ti\ ,\ tj,\ tstart,\ tend \in T$ (19)

5. EVALUATION

This paper has presented abstractions of various data filtering in a form of parallel coordinated graph. For evaluation, a dataset used is from a research center called Green Energy Research Centre (GERC). The center is located at high educational institution MARA University of Technology in Malaysia [24].

There are 11150 set of data which are collected every five minutes for the duration of 39 daysfrom various types of logs sensors. The sensors used are complied to international standard as a mean of quality control[25]. A total of nine types of variables have been collected from these logs sensors which are date, time, temperature, solar irradiance, ambient temperature, relative humidity, module temperature and wind speed. The dataset is complex enough to be used to generate abstractions of various visual filtering designs. It has variety type of variables; high volume of data and velocity type of data which are changing with time.

In this case study, C = {*date, time, solar radiation, gust speed, wind direction, ambient temperature, relative humidity, module temperature and wind speed*}.

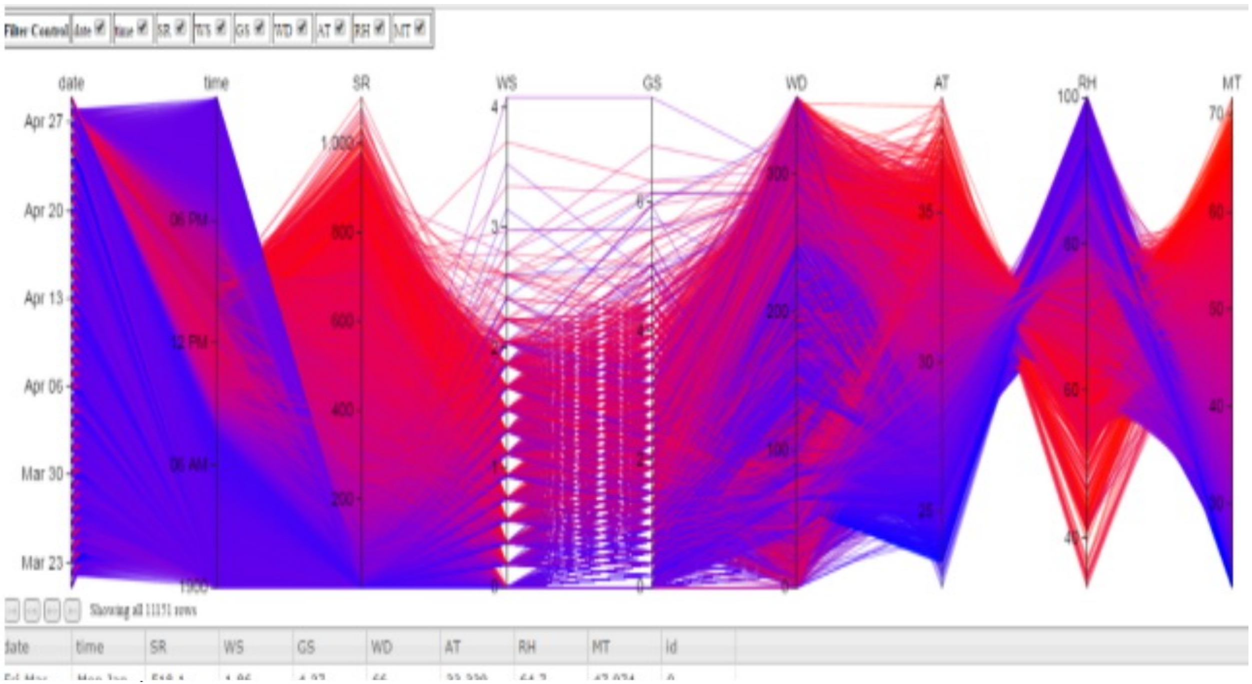

Fig.1. The solar photovoltaic data at the overview level

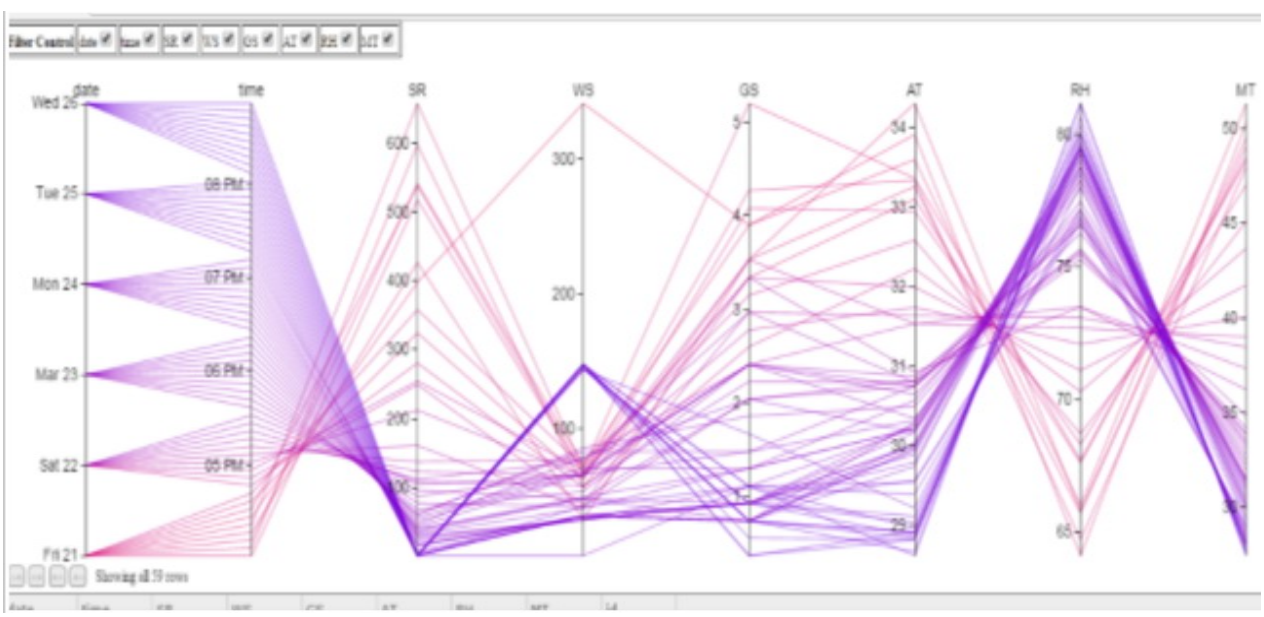

Fig.2. A filtered version of the solar photovoltaic data

All the seven filtering designed has been successfully implemented into the parallel coordinate graphs.

6. CONCLUSION

The findings support the overview of a multivariate data as well as drill-down capabilities through filtering in parallel coordinate graph.All the new variations in filtering have been embedded into the graph and have successfully support the various need of multivariate data. This research can be extended to support a wider visual analytical support such as zoom, brushing and sorting.

7. ACKNOWLEDGEMENTS

The authors would like to thank Universiti Teknologi MARA, Malaysia and Ministry of Higher Education Malaysia for the facilities and financial support under the national grant FRGS/1/2017/1CT04/UITM/02/2.

8. REFERENCES

[1] Stefanowski J. Data visualization: Or graphical data presentation. 2013. http://www.cs.put.poznan.pl/jstefanowski/sed/DM14-visualisation.pdf

[2] Biswas A, Dutta S, Shen H W, Woodring J. An information-aware framework for exploring multivariate data sets.IEEE Transactions on Visualization and Computer Graphics, 2013, 19(12):83-92

[3] Borup R. Data visualization for the database developer. 2015. http://www.ita-software.com/papers/Borup_DataVisualization_Published.pdf

[4] Wang L, Wang G, Alexander C A. Big data and visualization: Methods, challenges and technology progress.Digital Technologies, 2015, 1(1):33-38

[5] Andrienko N, Andrienko G.Visual analytics of movement: An overview of methods, tools and procedures. Information Visualization, 2013, 12(1):3-24

[6] Thomas J. J., Cook K. A. Illuminating thepath: The research and development agenda for visual analytics.New York: IEEE Press, 2005.

[7] Idrus Z, Abidin S Z Z, Omar N, Idrus Z, Sofee N S A M. Geovisualization of non-resident students' tabulation using line clustering. In Regional Conference on Sciences, Technology and Social Sciences, 2016, pp. 251

[8] Zhou H, Yuan X, Qu H, Cui W, Chen B. Visual clustering in parallel coordinates.Computer Graphics Forum, 2008, 27(3):1047-1054

[9] Glendenning K, Wischgoll T, Harris J, Vickery R, Blaha L. Parameter space visualization for large-scale datasets using parallel coordinate plots. Journal of Imaging Science and Technology, 2016, 60(1):104061-104068

[10] Heer J, Shneiderman B, Park C. A taxonomy of tools that support the fluent and flexible use of visualizations. Interactive Dynamics for Visual Analysis, 2012, 10(1):1-26

[11] Van Den Elzen S, Van Wijk J J. Multivariate network exploration and presentation: From detail to overview via selections and aggregations. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(12):2310-2319

[12] Luo W, MacEachren A M. Geo-social visual analytics. Journal of Spatial Information Science, 2014, 8(8):27-66

[13] Khan N, Yaqoob I, Abaker I, Hashem T, Inayat Z, Kamaleldin W, Ali M, Alam M, Shiraz M, Gani A. Big data: Survey,technologies, opportunities, and challenges. Scientific World Journal, 2014, 1-18

[14] ShneidermanB. The eyes have it: A task by data type taxonomy for informationvisualizations.In IEEE Symposium on Visual Languages, 1996, pp. 336-343

[15] Wang W B, Huang M L, Lu L F, Zhang J. Improving performance of forensics investigation with parallel coordinates visual analytics. In 17th IEEE International Conference on Computational Science and Engineering, 2014, pp. 1838-1843

[16] Idrus Z, Razak N H A, Talib N H A, Tajuddin T. Using three layer model (TLM) in web form design: WeFDeC checklist development.In 2nd International Conference on Computer Engineering and Applications, 2010, pp. 385-389

[17] Raidou R G, Eisemann M, Breeuwer M, Eisemann E, Vilanova A. Orientation-enhanced parallel coordinate plots. IEEE Transactions on Visualization and Computer Graphics, 2016, 22(1):589-598

[18] Sacco D, Motta G, You L I, Ma T Y. Chapter 5: Smart cities, urban sensing, and big data: Mining geo-location in social networks. In X. Liu, R. Anand, G. Xiong, X. Shang&X. Liu(Eds.), Big data and smart service systems. New York: Academic Press, 2016, pp. 59-84

[19] Shiravi H, Shiravi A, Ghorbani AA. A survey of visualization systems for network security. IEEE Transactions on Visualization and Computer Graphics, 2012, 18(8):1313-1329

[20] Hilda J J, Srimathi C, Bonthu B. A review on the development of big data analytics and effective data visualization techniques in the context of massive and multidimensional data.Indian

Journal of Science and Technology, 2016, 9(27):1-13.

[21] Kravetz S, Katz S. Goal directed approach to training parents of children with a developmental disability.Biritish Journal of Mental Subnormality, 1990, 36(70):17-29

[22] Omar N. Modelling complexities of learner's in handling web texts via abstract scene analysis. Malaysian Journal of Computing, 2014, 2(1):13-26

[23] Idrus Z, Abidin S Z Z, Hashim R, Omar N. Social awareness: The power of digital elements in collaborative environment.WSEAS Transactions on Computers, 2010, 9(6):644-653

[24] Green Energy Research Centre.Home. 2014, http://gerc.uitm.edu.my/gerc/gerc.php?site=home

[25] International Electrotechnical Commission. Photovoltaic system performance monitoring-Guidelines for measurement, data exchange and analysis. Geneva: International Standard IEC 61724, 1998

How to cite this article:
Idrus Z, Zainuddin H, Ja'afar A D M. Visual analytics: designing flexible filtering in parallel coordinate graph. J. Fundam. Appl. Sci., 2017, *9(5S), 23-32.*