Parallel Coordinates for Visual Analytics of Vehicle Navigation Area

Renzhi Zhang[1] and Peipeng Zhang[2]

[1]*College of Information Engineering, Huanghuai University, Henan, China*
[2]*Jiyuan vocational and technical college, Henan*
Corresponding E-mail: Renzhizz@126.com

Abstract

In this paper, we present a parallel coordinates of visual analytics technique, by taking advantage of parallel coordinates and time series plots. An application of its performance is proved on datasets associated with vehicle. Its performance include data selection, dimension transfer, color transform, plot inter-conversion, dimension zoom, dimension control of combination of parallel coordinates, and replace, display, append of time series plots. Furthermore, parallel coordinates of visual analytics can visually describe the similarity dimensions datasets associated with vehicle navigation area, and time series plots may be as an auxiliary means to further promote understanding and analysis of the correlations between involving vehicle properties, as well as it can visually depict any of two properties, specifically, with time as the horizontal axis, It is demonstrated that these measures present a promoting visual understanding for user to find an important attribute of vehicles, involving possible accidents in a certain period or stage.

Keywords: visual analytics, modularity parallel coordinates, time series plots, vehicle navigation area

1. Introduction

In the field of vehicle navigation study, the users are interested in discovering surrounding environmental factors or relationship between factors that can give rise to potential risks. However, high-dimensional pattern data is one of the vital performance data of the vehicle. So far, there is no absolute method to visual high-dimensional or/and multivariate data in inland vehicle system. The existing proposed techniques and tools of visualization of vehicle system mostly depend on human perception and experiences to gain insight. Generally, most of them demand human to analyzing and perceiving abnormal behaviors and risks on vehicle. In order to enhance the human understanding and perception of different levels of behaviors and risks of inland vehicle, visual analytics for vehicle has become a fervent research field in recent years that attempts to accelerate to understand and perceive surrounding environmental associated with vehicle in the processing of navigation.

Parallel Coordinates [1] have a distinguished and widely used technology for high-dimensional and multivariate data in analysis domain of visualization, that is first come up with by Inselberg [2], and a data element can be transformed in an N-dimensional space into a compact 2-dimensional visual representation by a polyline crossing parallel axes of each dimension for a multitude of multidimensional data sets. It or its variation has been applied to visualization, data mining, process control, and so forth, because it can concisely and switch display multidimensional data. Inselberg *et al.* [3] first put into use parallel coordinates method to solve the visualization problem, afterwards, some previous improvements have well done in regard to parallel coordinates method, such as curve instead of straight line enhanced visual effects [4], different hierarchy parallel coordinates displayed views[5], *etc.*

Visual analytics [6] is the vibrant and multidisciplinary research field of analytic reasoning facilitated by combining interactive visualization interfaces for devoting to human decision-making and knowledge discovery on the basis of very large and complex data sets. Its goal of facilitating the user to acquire a qualitative understanding of the information. A wonderful view of visual analytics distinctly indicates feature or structure within the data and then can help the user to preferably recognize patterns or detect abnormalities, or others, at the same time, without reducing information content or interfacing data in any way.

In this work, we propose a more novel design, visual analytics methods of modularity-based parallel coordinates (MBPC) to vehicle behavior and navigation area in inland river that contribute to users accurately get hold of changing trend of information in vehicle navigation area. Duo to in the process of the vehicle sailing, influenced by many environmental factors navigable waters (Figure 1). Therefore, it is particularly important to ascertain the connections between what factor or factors that have a significant impact on the ship, especially with potentially dangerous vehicles which may occur. This wok focuses on the aspects of vehicle order, vehicle availability, waterway, hydrology, weather, and its visual analytics. The results suggest that MBPC can be used as an alternative method for finding connections among a set of variables associated with vehicle and to enhance to understand environmental factors related to the vehicle around.
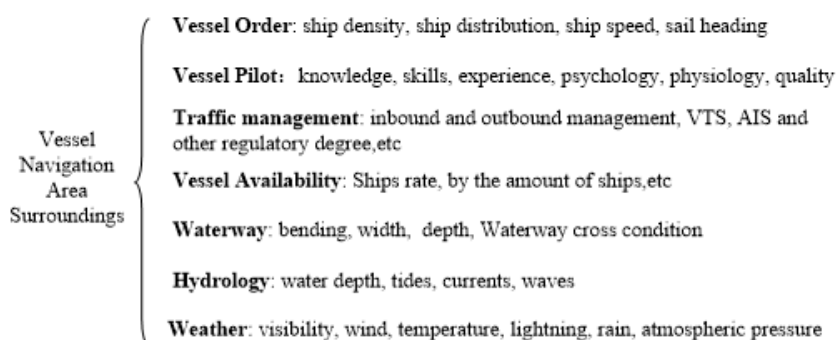

Figure 1. Vessel Navigation Area Surroundings

The contribution of this study and the specific benefits of MBPC are as follows:

- Cosine-based vehicle Properties Similarity Multidimensional Scaling(CVPSMS): Apache Spark[7], an emerging in-memory processing framework, has been facilitated Multidimensional Scaling algorithm.
- Modularity-based parallel coordinates (MBPC) integrates parallel coordinates sets with improving parallel coordinates for expression visualization surroundings of vehicle behavior and navigation area.
- Time series plots matrix display that row or column of 2-variable plots presented by each data element of adjacent points are connected with line segments.

The remainder of this paper is organized as follows. In Section 2 we give a review of related work. An overview of our system is described in Section 3. Followed by the implementation details of our techniques in Section 4. Finally, conclusions and suggestions for future work are presented in Section 5.

2. Related Works

To convey visual vehicle navigation area surroundings, it is typically studied in line with three steps in the processing of visual analytics: data collection and pre-processing, visual mapping and visual presentation, especially focus on in the last two steps.

2.1. Parallel Coordinates

This portion mainly contains two portions of parallel coordinates sets and improving parallel coordinates. Parallel coordinates is a technique pioneered by Inselberg [2]that has been applied for high dimensional analysis problems. Subsequently, it is used to the aspects of data visualization.

2.1.1 Similarity Multidimensional Scaling

Some research studies on similarity multidimensional scaling in visualization has been studied. Ankerst *et al.* [8] applied Euclidean-based distance with heuristic algorithms to search for similarity measure and optimal dimension order, and proved that arrangement problem of Euclidean-based distance was NP-complete. Dimension ordering algorithms in XmdvTool [9], users could manually vary the order of dimensions from a reconfigurable list of dimensions. Wei Peng *et al.* [10] presented dimension ordering algorithms to search for minimizing the visual clutter. These approaches mainly contributions lie in the ordering of axes, and to find axes layouts in parallel coordinator plots. Yang presented a hierarchical structure by grouping a host of dimensions so that reduced the complexity of the reordering dimensions. However, all these methods lack correlation expression between coordinates axes, at the same time, also unavoidably cause information loss, such as some information in the original data space may be lost, in one way or another. On the other hand, Euclidean-based distance metric units would be affected by index in different scale, it is generally required to be standardized, while the greater the distance, the greater the differences between individuals; in our work, we introduce space vector cosine similarity measure is not affected by the angle indicator scale, the value of cosine fall in the range [-1, 1], the larger the value, the smaller the difference.

2.1.2 Coordinate Axis Diversification

Few approaches have been done for extensions of coordinate axes. Fanea *et al.* [16] proposed parallel glyphs that coordinate axis had been extended to star glyphs to facilitate comparison of the data and provide interactive However, its methods only the visual comparison of non-adjacent items, at the same time, data set is not to large and not easy to observe for non-professional visual users. Mao Lin Huang *et al.* proposed an arc-based coordinate plots to identify the features of network attacks by displaying visual patterns, because the length of arc is longer than the line segments, the density of points displayed in each axis could be enlarged in in parallel coordinates plane. However, it lacks some interaction techniques except for brush and hardly figuring out the relationship between attributes in non-adjacent positions. The curves possessing some statistical property linking data points on adjacent axes are described in literature, it utilize the splatting framework to detect clusters and reduce visual clutter. These approaches of corresponding extensions of the axes mostly and still focus on the line segments, and apply curves to joint the vertices between the two adjacent axes.

In this work, we take full advantage of merit of parallel coordinates sets and make better generalized parallel coordinates to offset the disadvantages they have separately, so that deliver and process multidimensional data associated with the vehicle behavior.

2.1.3. Time Series Plots Matrix

Time series plots are frequently visualized using line plots, where line segment denote the change of data points over time. It consists of a row or column of 2-variable plots with a common axis, particularly a time variable. By default, each data element of adjacent points are connected with line segments. Zhao *et al.* present circular temporal histograms to express time series so that explore the dependency of movement behaviors, a time

graph or temporal histogram are proposed respectively to combine density maps with time series displays. These approach of coloring and shading of ring segments show aggregate values, where the bigger the number of the space compartments and the longer time series, it may be difficult to detect the spatio-temporal distribution of object presence. On the other hand, data associated with the vehicle may change or not be uniform in structure, each factor may have special characteristics, and measurements on different variables may be taken each time. These irregularities yield data that is difficult to work with. To analyze these data, we denote the response variables to be as a vertical axis, by default, and the time-dependent explanatory variables to be as a horizontal axis.

Eventually, we test our approaches in data sets associated with vehicles and demonstrate that these interactive approaches are effective in revealing the connections between different axes, at the same time, easier and deeper to understand the results of visualization and analyze vehicle navigation area surroundings.

2.1.4. Data and Pre-processing

The data used in this study is the data of Automatic Identification System, Vessel Traffic Services, Closed Circuit Television, Global Positioning System of over thousands of vessels in Yangtze River in Hubei Province, China, especially focus on three areas of Wuhan city, Huanggang city, Ezhou city. Most of data record mainly concludes static data (*e.g.* name and type of vessel, *etc.*), dynamic data (*e.g.* geographic longitude and latitude of each vessel, *etc.*), navigation data (*e.g.* status, destination, *etc.*), territorial waters traffic environment (*e.g.* weather, water flows, *etc.*), and so forth.

On account of the uncertainty of data transmission and data structure may be inconsistent, the relevant semantic information defined and stored repeatedly. Therefore, in order to the accuracy of the experiment, we adopt the following rules for cleaning and repair:

(1) using the mean method for missing a few items in a col. or row in database, and deleting some rows or cols if the data of this row or col in database miss exceed 60%;

(2) employing level quantify for non-numerical values, such as the type or level of accident;

(3) utilizing artificial and historical experience to repair some error data.

3. Implementation Details

All visualization algorithms and performance are implemented and measured for our study with experiments on a ThinkCentre desktop with inter(R) core(TM) i7-3770 CPUs 3.4 GHz and 8 GB Memory. The graphics card used is NVIDIA GeForce GT 620 with 1024M of SDDR 3 Memory. The software environment is based on CentOS with VTK.

In the following subsections, implementation details about spark facilitated CSMDS and more specification on system interface and interaction for users will be discussed.

Cosine-based Vehicle Properties Similarity Multidimensional Scaling(CVPSMS)

A different dimension axes ordered in the display will reveal different relationships of the data, and affect the user data analysis and perceived judgment for visualization of multidimensional data. An intuitively efficient of relevant properties and abnormal data have been found for users by the favorable order of dimensions.

In the our module part of improving parallel coordinators, the property dimensions axis reordered is handled, and to remedy this shortcoming of current tools for vehicle itself information and inland navigation river of relational similarity between neighboring coordinates axes by Cosine-based method in the following procedures.

Let D be the sets of N-dimensional object (as follows(1-4)):

$$D = \{X_1, X_2, \ldots, X_i, \ldots, X_j, \ldots X_n\} \tag{1}$$

Whereby:

$$X_i = ((x_i1, x_i2, \ldots, x_ik, \ldots, x_im)^T \mid 1 \leq i \leq N, 1 \leq k \leq m) \tag{2}$$

$$X_j = ((x_j1, x_j2, \ldots, x_jk, \ldots, x_jm)^T \mid 1 \leq j \leq N, 1 \leq k \leq m) \tag{3}$$

where m is the number of X_i or X_j data items, and N is the dimension of the data, and x_{ik} or x_{jk} is denoted as k-th property value in i-th or j-th respectively. $sim(\theta_{ij})$ is presented with relational between i-th and j-th to determine the similarity condition:

$$sim\theta_{ij} = \cos(X_i, X_j) = \frac{\sum_{k=1}^{m}(x_{ik} \times x_{jk})}{\sqrt{\sum_{k=1}^{m}(x_{ik})^2} \times \sqrt{\sum_{k=1}^{m}(x_{jk})^2}} = \frac{X_i \cdot X_j}{|X_i| \times |X_j|} \tag{4}$$

The larger the value of $sim(\theta_{ij})$, the higher the similarity between axis. Usually, the most similarity between X_i or X_j describes close to distribution trends in these two properties axes in vehicles multi-dimension data sets.

Spark Facilitated CVPSMD

To deal with large amounts of movement data, some spatio-temporal data can be done within a data warehouse or a database, spatio-temporal data here are loaded in RAM for visualization and interactive analysis. Borg. Proposed a fundamental theory and applications of modern multidimensional scaling that is a means of visualizing the level of similarity of individual cases of a data sets. Ingram *et al.*. presented a multi-level multidimensional scaling algorithm based on a parallel force-based subsystem simulation to exploit graphics processing unit (GPU) hardware while the GPU parallelism improved speed of computation. However, these approaches either apply conventional data read mode, or load the entire data sets，it may suffers from extremely serious computational and time costs. In our work, CVPSMD is implemented by Apache Spark, an emerging in-memory processing framework, to facilitate similarity multidimensional scaling.

The approach we present in this work is to add or subtract dimension into cosine similarity multidimensional scaling configuration, where is firstly constructed and initialized. Secondly, when completing initialized cosine dimension ordering configuration, CVPSMD method is introduced into spark to facilitate dimension ordering. Next, the users can drag axes into the plot, some new dimensions are added or subtracted to calculate similarity among different neighboring axes in the visualization process of parallel coordinators.

Table 1. Performance Comparison of CPU-based, GPU-based, Spark-based

Data	Dimensions	Data Size	CPU-MS	CPU-SimMS	GPU-MS	GPU-SimMS	Spark-MS	Spark-Euclidian-MS	Spark-SimMS
Car	8	456	3724ms	823ms	728ms	425ms	326ms	215ms	93ms
Vessel	24	2167	13284ms	9284ms	5425ms	2936ms	1053ms	842ms	604ms

Comparing the performance here CPU-based and GPU-based with MS and Similarity MS (Sim-MS),and Spark-based with MS, Euclidean MS and Sim-MS to facilitate multidimensional scaling. By testing visual approach with two datasets of car and vehicle with different sizes, and compare them and with our observation, as showed in Table 1,Spark-based Sim-MS improves the performance over previous other methods, especially,

face with when the data dimensions is progressively increasing. In our implementation, the most common situation is that the user add dimensions among axes into the plots step by step, the results display that Spark-based Sim-MS comes observably about computing improvements. The velocity of time cost changes from 823ms to 93ms and from 9284ms to 604ms by CPU-SimMS and Spark SimMS respectively, relying on the data dimensions with size. Spark facilitated CVPSMD can offer about 10 times improvement.

Representing Vehicle Properties Coordinate Axis Diversification

Few approaches have been done for extensions of coordinate axes. Some of techniques adopt in the form of a single extension of axis or the strategy of mapping each single data into a coordinate axis display on the screen. We employ the mixture of three different coordinate axis that can be adjusted according to specific requirements.

In this paper，we present a novel coordinates axes, which consist of diversification to extend representation of the coordinate axes in integrated space of parallel coordinates visualization. We should touch on, that the frequency-based visualization method for contribution rate of each axis attribute to dynamic data and navigation data.

The first type of axis is based on user-defined standardization(USVA) have been proposed. The axis can be standardized as the uniform forms and designed based on user-defined requirement, as follows(6),

$$\frac{v'-L'}{L-L'} = \frac{v - \min(x_i)}{\max(x_i) - \min(x_i)}$$

(6)

Where $\max(x_i)$ and $\min(x_i)$ are the maximum and minimum values for each axis of the original data respectively, v is the value of an axis, L and L' are the maximum and minimum values for each axis after standardization respectively, usually let L' is as 0 or user-defined, let L as 1 or user-defined, suggest as a positive integer. v' denotes corresponding to v values for each axis after standardization.

The second types of axis as bar-based (BBA) have been presented. the axis was partitioned into k sub-section by the bars, l denotes the length of coordinate axis of $i-th$ attribute axis, m denotes the total number of data items of $i-th$ attribute axis.

Each bar has height $H(k_i) = Num(k_i)/m$, $Num(k_i)$ denotes the number of data items of $i-th$ sub-section in the k section.

Each bar has width $W(ki) \leq d$;

Step1: if $d_{pq} \leq d$,then $W(k_i) = d_{pq}$,go to step2;

Step2: p-1 and if $d_{p-1,p} \leq d, d_{p-1,p} + d_{pq} \leq d$,then $W(k_i) = d_{p-1,q}$,go to step3;

Step3: q+1 and if $d_{q,q+1} \leq d, d_{p-1,q} + d_{q,q+1} \leq d$,then $W(k_i) = d_{p-1,q+1}$,go to step2;

Step4: if $d_{pq} \geq d$,then $W(k_i) = d$.

$$d = \frac{d_{12} + ... + d_{pq} + ... + d_{m-1,m}}{l}$$ is a reference value and a maximum width

value, d_{pq} is distance between any two adjacent points of $i-th$ attribute axis, the sets of d_{pq} as $\{d_{12}, d_{23}, ... d_{pq}, .. d_{m-1,m}\}$.

These bars may be viewed when the mouse hover the corresponding to axis, meanwhile, the size of data associated with bar will also be showed, furthermore, when double-click sub-section bar, corresponding to data items can be higher highlight in the display. For each bar, the count of number has been computed from the counts contained in the bar.

The thirdly types of axis as layer-based (LBA)have been proposed, which was divided into several section according to specific requirements, it similar to Parallel Sets proposed by Robert Kosara, such as concluding three sub-section of Wuhan city, Huanggang city, Ezhou city. Nevertheless, layer-based axis can support numerical data statistical classification, but not in the literature.

The three different types of axis are seamlessly integrated and can be adjusted by some interactive techniques according to user-controlled requirements. Parallel coordinates are displayed by controlling file menu button. We see the purple cures that are error value in the datasets of vehicle, not outliers. Statistical classification of data are processed, the colors are randomly added in the BBA. Users can choose their own color in the LBA or USVA, and can also for a variety of performance operation, such as dimension zoom, dimension control, *etc.* Brushing in after an operation, such as brush in red, the mouse is placed on point on the axis.

Representing any Two Involving Vehicle Properties

Time series are special in that time points where the patterns of data have a relatively fixed order. Along with the interactive visualization interface, our system provides temporal-based views that combined with other properties, such as vehicle speed, water flow, interacting with parallel coordinates.as shown in the low right part Figure2. Time series plots matrix is displayed by controlling file menu button. Note, to reduce clutter point of time for the rounding process. For example, 13:20 becomes 13.
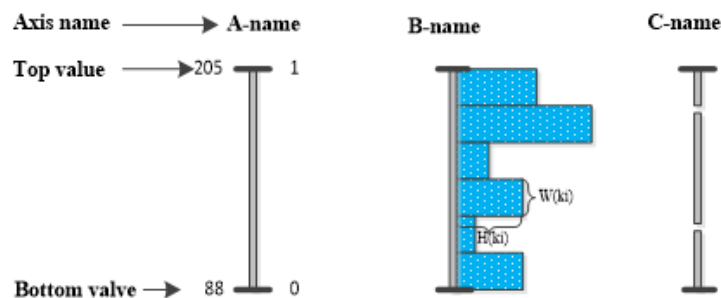
(a) User-defined Value Axis (USVA) (b) Bar-based Axis (BBA) (c) Layer-based Axis (LBA)

Figure 2. Parallel Coordinates

The line graphical visualization allow users to overlay different features by using check box provided against time feature for easy comparison and to observe visual trends. Users can select on one or quite a few accident cases handled by Yangtze River police stations and comprehend information at the time of the accident, such as wind power. The number of accident cases can be easily observed the peaks in the summer months for the middle reaches of Yangtze River, enabling users to detect the seasonal trends of their data sets and giving them a reference information and helpful precautions in their own voyage. The time series display indicates mostly regular character of the movement of the vehicles. The variation of the vehicle presence can be analyzed and explained by navigation condition, such as weather conditions. This system also allows users to weekday and weekly trends of accident cases, and further gives assistance for users to acquire an more effective vehicle behavior scheme.

Representing any Two Involving Vessel Properties

Time series are special in that time points where the patterns of data have a relatively fixed order. Along with the interactive visualization interface, our system provides temporal-based views that combined with other properties, such as vessel speed, water

flow, interacting with parallel coordinates.as shown in the low right part Figure2. Time series plots matrix is displayed by controlling file menu button. Note, to reduce clutter point of time for the rounding process. For example, 13:20 becomes 13.

The line graphical visualization allow users to overlay different features by using check box provided against time feature for easy comparison and to observe visual trends. Users can select on one or quite a few accident cases handled by Yangtze River police stations and comprehend information at the time of the accident, such as wind power. The number of accident cases can be easily observed the peaks in the summer months for the middle reaches of Yangtze River, enabling users to detect the seasonal trends of their data sets and giving them a reference information and helpful precautions in their own voyage. The time series display indicates mostly regular character of the movement of the vessels. The variation of the vessel presence can be analyzed and explained by navigation condition, such as weather conditions. This system also allows users to weekday and weekly trends of accident cases, and further gives assistance for users to acquire an more effective vessel behavior scheme.

Conclusions and Future Work

In this paper, we present a modularity-based parallel coordinates of visual analytics technique, by taking advantage of parallel coordinates and time series plots. Its performance demonstrates that modularity-based parallel coordinates are not only closely related each other but also seamlessly integrated together. These measures provide a promoting visual understanding for user to find an important attribute of vehicles, involving possible accidents in a certain period or stage. In the future, we will consider the factors of the vehicle pilot or vehicle traffic management that are added to future work, due to incomplete data. Simultaneously, we will also consider human factors, and its vehicle visual analytics, in order to further facilitate accurate visualization of the vehicle.

References

[1]   A. Inselberg and B. Dimsdale, "Parallel coordinates: a tool for visualizing multi-dimensional geometry", In Proceeding of IEEE Visualization, **(1990)**, pp. 361-378.
[2]   A. Inselberg, "The plane with parallel coordinates", The Visual Computer, vol. 1, no. 2, **(1985)**, pp. 69-91.
[3]   A. Inselberg and B. Dimsdale, "Parallel coordinate, Human-Machine Interactive system", Berlin: Springer, **(1991)**, pp. 199-233.
[4]   M. Graham and J. Kennedy, "Using curves enhance parallel coordinate visualizations", Proceedings the seventh International Conference on Information Visualization. Washington DC: IEEE Computer Society, **(2003)**, pp. 10-16.
[5]   Y. H. Fua, M. O. Ward and E. A. Rundenteiner, "Hierarchical parallel coordinates for exploration of large datesets", Proceedings of the Conference on Visualization 99: Celebrating Ten Years. Los Alamintos: IEEE Computes Society Press, **(1999)**, pp. 443-508.
[6]   D. Keim, G. Andrienko, J. D. Fekete, C. Gorg, J. Kohlhammer and G. Melanc, "Visual Analytics: Definition, Process, and Challenges", Information Visualization, **(2008)**, pp. 154-175.
[7]   M. Zaharia, M. Chowdhury, M. Franklin, S. Shenker and I. Stoica, "Spark: Cluster computing with working sets", HotCloud, **(2010)**.
[8]   M. Ankerst, S. Berchtold, and D. A. Keim, "Similarity clustering of dimensions for an enhanced visualization of multidimensional data", Proceeding IEEE Symposium on Information Visualization, **(1998)**, pp. 52-60.
[9]   Xmdvtool home page. http://davis.wpi.edu/ xmdv/.http://davis.wpi.edu/~xmdv.
[10]   W. Peng, M. O. Ward and E. A. Rundensteiner, "Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering", IEEE Symposium on Information Visualization, no. 15, **(2004)**, pp. 89-96.

Author

Renzhi Zhang, assisted professor, Research direction: computer navigation.