

A Visual Analysis Tool That Smoothly Switches Between Tabular Forms and Parallel Coordinates

Kazuo Misue
Department of Computer Science
University of Tsukuba
misue@cs.tsukuba.ac.jp

Takashi Yuki
Department of Computer Science
University of Tsukuba

ABSTRACT

An intermediate representation between tabular forms and parallel coordinates is proposed. The new representation makes it possible to switch back and forth between tables and parallel coordinates in smooth transformations. The authors have developed a tool that can manipulate these representations for the analysis of multidimensional data. Through experimental evaluation, the advantages of this tool are illustrated. It is expected that both the representation and the tool offer a significant improvement to the early stage of data analysis.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interface—*graphical user interface, interaction styles*

General Terms

Design, Human Factors

Keywords

Information visualization, visual analysis, multidimensional data, parallel coordinates, tabular form

1. INTRODUCTION

A variety of visualization methods are used in data analysis processes. Tabular forms are often used as the fundamental representations in the early stage of data analysis. Parallel coordinates[3] are one of the most widely used ways of visualizing multidimensional data. That is, data in a table are commonly translated into parallel coordinates during the course of an analysis.

The purpose of our study is to make analytic tasks that involve multidimensional data more effective. Our premise is that combining a tabular form with parallel coordinates would be useful if a translation from tables into parallel coordinates is performed.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

VINCI 2011, August 4-5, 2011, Hong Kong, China.
Copyright 2011 ACM 978-1-4503-0875-5/11/08 ...\$10.00.

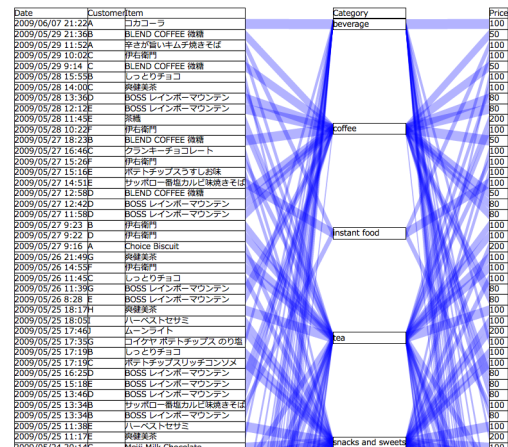


Figure 1: Example of intermediate representation between tabular forms and parallel coordinates.

The initial challenge, however, is that they are quite different representations, which can make it difficult for a given representation to play a role as a table and also as parallel coordinates.

We, therefore, propose a process of switching between a tabular form and their corresponding parallel coordinates in smooth transformations that preserve their common characteristics. To this end, we developed an intermediate representation between a tabular form and parallel coordinates (see Figure 1). These two representations can smoothly switch back and forth between each other via the intermediate representation. This paper describes that intermediate representation as well as the tool that manages this task.

2. BACKGROUND

2.1 Data

The object data of the analysis that we assume is multidimensional data. The multidimensional data consists of one or more, but usually many, data sets. Each data set has n values ($n \gg 1$).

2.2 Table

In a table, in most cases, each value is expressed by a character string in one cell, and one data set is expressed in a row of cells on a horizontal line. The values in each dimension are arranged vertically. We can traverse the values of

a given dimension by following the cells vertically. We may also change the order of the dimensions appropriately.

The fundamental model of tables is based upon two-dimensional matrix placement. This model enables many people to understand data without the need for a detailed explanation.

2.3 Parallel Coordinates

In a set of parallel coordinates, each dimension is expressed on a vertical axis. Each value is expressed at the position of a point on an axis, and one data set is expressed by connecting the points on all axes using line segments. We can grasp the distribution of the values of a dimension by observing the placement of the points on the axis. We may also appropriately change the order of axes that correspond to the dimensions.

2.4 Comparison between Tabular Form and Parallel Coordinates

In tables, values are expressed by character strings that make it possible for us to precisely read the values in detail. It is not easy, however, to read an overview such as the distribution of values in this way. For this reason, a table is unsuitable for an analysis based on an overview of values.

Parallel coordinates visually represent the distribution of values and the correlation between dimensions so we can intuitively gain an overview of values. That said, it is not easy to read every value precisely. Detail data may be hidden by other data depending on the distribution of values.

Both tables and parallel coordinates are widely used representations in data analysis. They have different features, and the drawbacks of one are compensated for by the advantages of the other.

3. RELATED WORKS

Visualization is a widely used and effective approach to data analysis. Many visual representations have been proposed that are in use. In this section, we introduce some important works related to table and parallel coordinates that are relevant to this study.

3.1 Extension of Table Representation

In the initial stage of most data analysis processes, a table is used as the format for the arrangement of data[1, Chapter 1]. In typical tables, character strings are arranged in cells and do not always assume a visual form of expression. However, tables do possess an aspect of UI for a large-scale database.

With the aim of improving the power of table expression, a number of studies have expanded tabular forms while maintaining the basic placement style. TableLens enables users to intuitively grasp large-scale multidimensional data sets by changing the sizes of cells according to the user's viewpoint and arranging charts in cells in order to visualize the values[7]. John et al. expressed data in TableLens by using a transparency[4], which made it possible to grasp more detailed information. Telea combined a table with a treemap[8], which made it possible to grasp the values and the hierarchical structure related to them. The arrangement of various representations in tabular form is another way of creating an extension of a table. A scatter plot matrix is an example of such an extension. Another example is matrix

representation of dynamic networks, in which a snap shot of a changing network is expressed in each cell.

Our focus here is on tables that represent each value in a cell. Our aim is to create a table that corresponds to a given set parallel coordinates by making use of the fact that tables are often used with parallel coordinates.

3.2 Extension of Parallel Coordinates

Various studies have been performed whose aim has been the extension of parallel coordinates. In some studies, parallel coordinates were extended with the principal aim of promoting the efficiency of multidimensional data analysis. Other studies developed representations that adopted selected features of parallel coordinates for a number of different purposes. Still others have tried to combine parallel coordinates with some other representations.

Yuan et al. combined parallel coordinates with scatter plots[9]. A scatter plot is drawn in any area between two axes. Parallel Sets is a method of visualization for categorical data[5]. This method is based on an axis layout of parallel coordinates, but shows data frequencies instead of the individual data points. Parallel Tag Clouds is a combination of parallel coordinates and tag clouds[2]. Tags are laid out on implicit axes and links are drawn as moderate lines, an approach that makes the comparison of sets of tags more easy. Lex et al. developed the Caleydo Matchmaker, which facilitates the comparative analysis of multidimensional, quantitative data[6]. It uses a parallel coordinates metaphor, and uses head maps rather than simple lines as axes.

Our aim here is to introduce the features of parallel coordinates into a table by making the arrangements of cells more flexible.

4. INTERMEDIATE REPRESENTATION

We designed an intermediate representation between tabular form and parallel coordinates. This representation can be partially in the table form and partially parallel coordinates. We can smoothly switch back and forth between visualizations in the two representations via the intermediate representation.

4.1 Design of Intermediate Representation

In a table, one of the most widely used ways is to express each data set by lining up cells horizontally. In this way, values in the same dimension are arranged vertically. In parallel coordinates, each value is expressed at the position of a point on a vertical axis. Both representations express a dimension in a vertical alignment. Our concept was to design an intermediate representation by focusing upon this common feature.

Expression of a Value.

In the same way as in an ordinal table, a single value is expressed by a character string in a rectangle (a cell).

Expression of a Dimension.

In the same way as in a table, a dimension is expressed by arranging cells vertically. No axes are explicitly drawn as parallel coordinates, though the result is the same as in parallel coordinates in that a dimension is expressed in a vertical arrangement. We can call such an arrangement a "virtual axis."

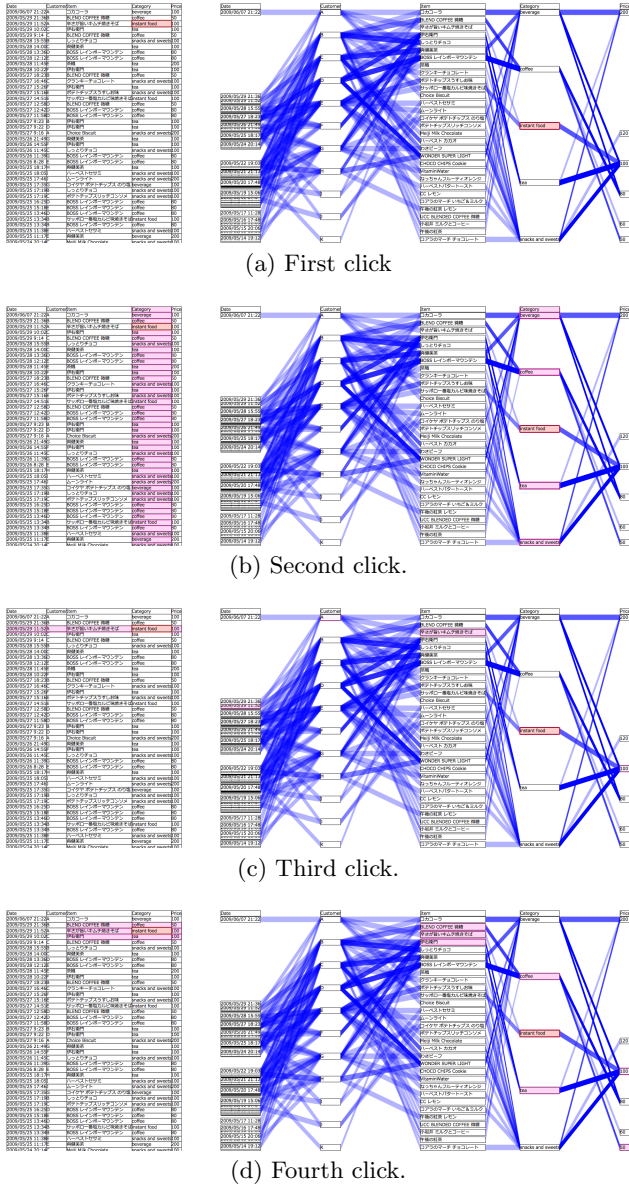


Figure 3: Logical selection. Figures on the left-hand side are in tabular forms and those on the right-hand side are in parallel coordinates.

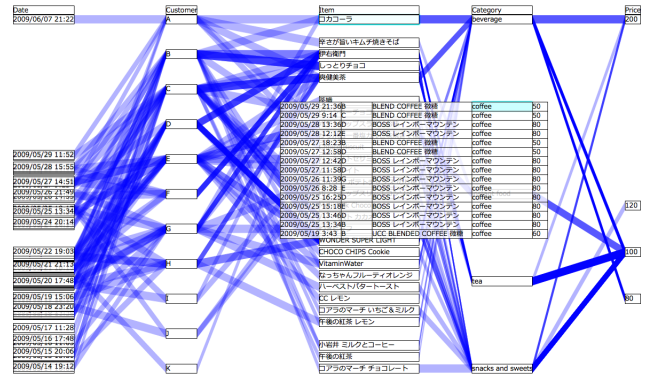


Figure 4: Example of a temporal table.

axes at the appropriate intervals (see Figure 2(b)). When we use this command, the transformations of the view are shown in animation.

Vertical Transformation.

We can use another command to change the positions of the cells on an axis (see Figure 2(c)) and yet use another command to change their positions on all axes (see Figure 2(d)). If we rearrange all the axes in appropriate positions and change the positions of all cells according to their values, we then obtain a set of parallel coordinates.

Temporal and Partial Table.

We can create a partial table by collecting data sets that have the same value as a specified cell. Related cells are made to come together to the specified cell via a hovering operation that is done with the mouse. Figure 4 is an example of such a representation that includes a temporal table. This function works as a tool tip, so it is quite easy to create a table in which we can see the original data set.

5.5 Grouping

Two adjacent dimensions can be grouped into one dimension, an operation that makes it possible to reduce the number of dimensions. When the positions of cells express their values, the position of a pair of grouped cells follows one in the pair.

5.6 Shifting

The Shift operation moves cells that are related to a specified cell (see Figure 5). It is similar to the logical selection with highlights given above, but it also changes the positions of cells. This provides more impressive and useful visual effects by combining the highlighting of cells.

Because the Shift operation creates a subset of cells, it works as a kind of filtering. When we want to focus upon a subset of cells with particular properties, we can then create that subset spatially.

6. USAGE EXAMPLE

Here we describe the tool being applied to a practical example.

6.1 Objective Data

Objective data was extracted from a sales database managed in our laboratory. While it is not the database of an

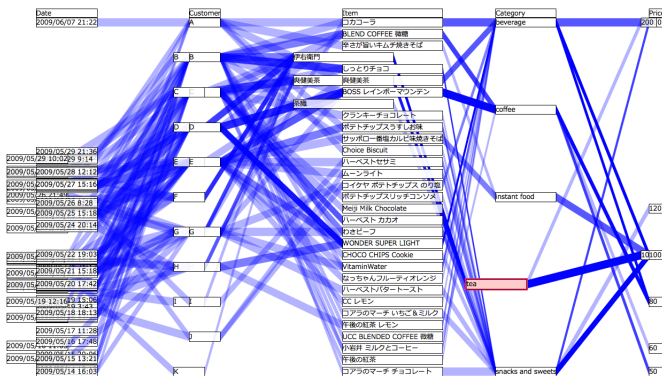


Figure 5: Cells related to *tea* has been moved using the Shift operation.

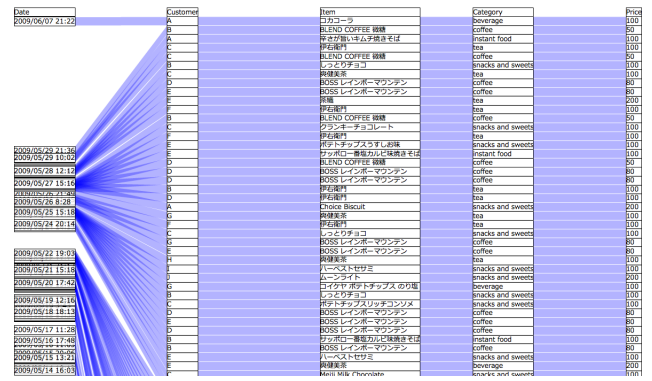


Figure 6: Values on the *time* axis have been rearranged.

actual shop, the transactions whose records were used are real.

Each record represents a single transaction that has six attributes: purchase date, time, customer name, item name, item category, and price. The customer name, item name, and item category are categorical data. The other three attributes are numerical data.

We used data that were recorded between May 22 and June 7, 2009. The number of records was one hundred, and these records represent ten different customers. In addition, the item name attribute included twenty-one different kinds of items.

6.2 Purpose of Analysis

The aim of the analysis was to gain insight into the purchasing tendencies of each customer.

6.3 Operations

After reading the data, the tool displays the data in a table. From that table, we understand that there are six kinds of attributes in the data, and we can read the values for each attribute.

First, we focused our attention on the *time* and then proceeded with the analysis. We changed an attribute of *time* into parallel coordinates in order to visualize the distribution of the times when items were purchased (see Figure 6). We were able to see that the distribution according to *time* was divided into a number of separated by several gaps. The small gaps indicate that items not purchased at midnight, and the large gap indicates a holiday. Such features cannot be readily ascertained from a table, but can be immediately found once the values are placed on a time axis. If we note some interesting features of the distributions, we can easily proceed with a more detailed analysis of those features.

We transformed all attributes into a set of parallel coordinates so that the analysis could expand to consider not only *time* but all the other attributes as well. We understood that *tea* and *coffee* were the predominant purchases due to the darkness of the line that could be seen when we focused upon *category*. We compared the purchase times of both categories in order to gain insight into the time periods in which they sold well. We obtained the time when tea sold well by using the Shift operation. When we selected the cell of category *tea* and moved it to the left, the cells of the purchase *time* of *tea* also moved to the left. We were then easily

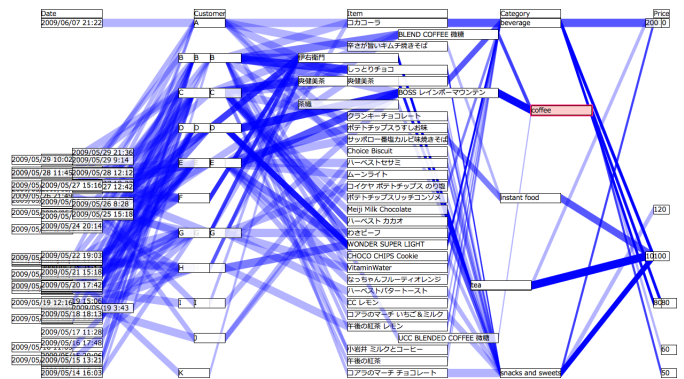


Figure 7: Cells related to *tea* and *coffee* have been moved using the Shift operation.

able to see that several bottles of tea had been purchased every day.

When we also moved *coffee* to the right-hand side, we found that coffee had not been purchased every day, but had been purchased in relatively greater quantity at some time (see Figure 7). We could easily compare the differences in the related data by performing the Shift operation and moving back and forth between these two categories.

The cells left on an axis after performing the Shift operation twice were of the data related to categories except for *tea* and *coffee*. In these data, we could then observe the distribution of the entire data set. This shows that we were able to gain insight into an overall set of information by focusing our attention only on certain subsets. For example, judging from the entire data set, many items seemed to be purchased every day. However, when we focused on the categories of coffee and tea, we found that only one item in the other categories had been purchased on each of the other days. In other words, we found that coffee and tea actually accounted for many of the sales in the entire data set.

7. EXPERIMENTAL EVALUATION

We performed an experimental examination to evaluate the effectiveness of the representation and the tool we developed. We performed the same tasks both with our tool and a comparison tool, and compared the results in terms of time and accuracy.

Table 1: Tasks performed in the examination.

No.	Questions ¹	Attr. ²	Cat. ³
#1	How much does the most expensive item cost?	1	s
#2	What is ranking of beverage on the basis of sales?	1	s
#3	What is the item whose sales ranking is 6?	1	s
#4	What is the most sold item among items priced at 200 yen?	2	s
#5	What is the most sold item in the category with a sales ranking of 3?	2	s
#6	Who bought the maximum number of <i>Harvest Cacao</i> ?	2	s
#7	Who bought the item that was sold the most among the lowest priced items?	3	s
#8	Among the people who purchased <i>Iyemon Tea</i> , when did the person who bought the fewest items make the purchase most?	3	s
#9	Who bought the maximum number of the most sold item in the most sold category?	3	s
#10	Between C and D, which was the one that bought items regularly?	2	c
#11	Between tea and beverage, which item was sold regularly?	2	c
#12	What is sales ranking of <i>BOSS Rainbow Mountain</i> among all items?	1	s
#13	When was the time when the first product was sold?	1	s
#14	What are the rankings of the items purchased by H (arrange in increasing order)?	1	s
#15	Who bought the maximum number of the items that had a sales ranking of 2?	2	s
#16	What is the most sold category in the purchase history of E?	2	s
#17	What is the most sold item in the category with a sales ranking of 2?	2	s
#18	Which item did the person who bought the maximum number of 200-yen items buy the most?	3	s
#19	Who bought the most items in the category to which the second-sold item belongs to?	3	s
#20	On which day was the most sold item in coffee sold most?	3	s
#21	Between E and F, which was the one that bought items regularly?	2	c
#22	Among the items priced at 200 yen and 80 yen, which items were sold regularly?	2	c

¹ All questions were in Japanese.

² The number of attributes required focus.

³ Category of tasks: “s” denotes value-specifying tasks and “c” denotes value-comparing tasks.

7.1 Comparison tool

We also prepared a tool to perform comparisons. This tool provides an exclusive view of each table along with a set of parallel coordinates. The tool has fundamental functions for each table and its parallel coordinates. Linking and brushing is provided to make the correspondence between the two views clearer.

7.2 Tasks

We prepared twenty two tasks that had varying degrees of difficulty and characteristics, but all of which had unambiguous answers. Table 1 shows questions in all twenty tasks. There were two kinds of tasks: value-specifying tasks and value-comparing tasks. An example of a value-specifying task would be to pose the question, “How much does the most expensive item cost?”; an example of value-comparing tasks would be the question “Between C and D, which was the one that bought items regularly?”

Performing a task requires focus upon at least one attribute. We varied the number of attributes receiving such focus from one to three so as to vary the degree of difficulty of the tasks. For example, if we can complete a task by focusing upon the price, the number of attributes of that task is one. If we must focus upon the price and the category, the number of attributes is two. The latter task is, of course, more difficult than the former.

7.3 Users

Our users were six undergraduate and graduate students majoring in computer science. Because they had no experience with either tool, they were allowed to practice using both tools prior to the experiment in order to become familiar with them.

Every user asked to perform eleven tasks with our tool and the other eleven tasks with the comparison tool. For a

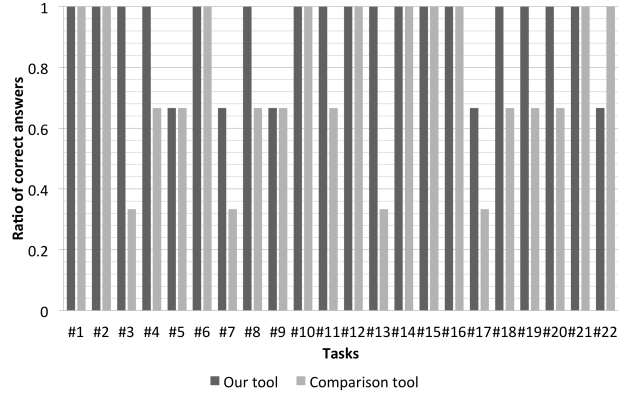


Figure 8: Ratio of users who answered correctly.

task, three users performed with our tool and the other three users with the comparison tool.

7.4 Results

First of all, we focused on how many users could answer correctly. Figure 8 shows the ratio of users who answered correctly. The ratio for the comparison tool was 0.33 for some tasks. The ratio for our tool was 0.66 or more for all tasks. The ratio for our tool was higher than or equal to the ratio for the comparison tool for all tasks, except for one. We attribute this difference to a users misunderstanding the correspondence between the two views in the comparison tool. Indeed, a few users gave up before completing some tasks with the comparison tool.

Figure 9 shows the average completion time for each task. For most tasks, the time taken with our tool was shorter than the time taken with the comparison tool. For some

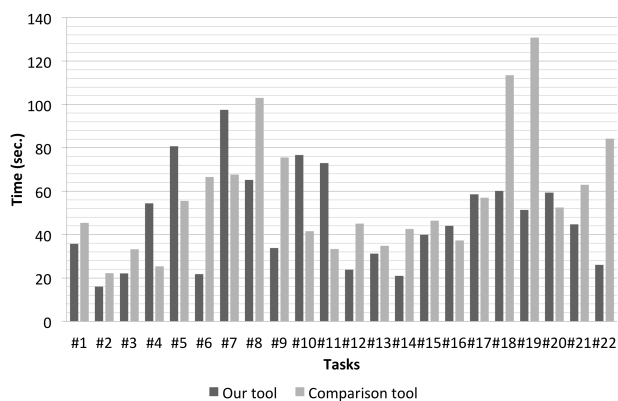


Figure 9: Task completion time.

tasks with two attributes, however, the reverse was true. Because our tool shows the transformations through animation, we surmise that the overhead associated with the animation resulted in a longer time for the tasks with two attributes. For more complex tasks with more attributes, the time taken in the case of our tool was shorter. The overhead may be ignored for such complex tasks.

8. CONCLUSION

In this paper, we have proposed an intermediate representation between tabular forms and parallel coordinates. This representation makes it possible to switch back and forth between these two representations in smooth transformations. We also developed a visual analysis tool that provides functions for the management of these representations. By performing a number of tasks that involved reading information from a set of multidimensional data, we confirmed that our tool offered advantages in terms of correctness and the time taken to perform tasks. We anticipate that this will make it possible to streamline and facilitate the early stage of the analysis process for multidimensional data.

9. REFERENCES

- [1] S. K. Card, J. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, 1999.
- [2] C. Collins, F. B. Viegas, and M. Wattenberg. Parallel tag clouds to explore and analyze faceted text corpora. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology, 2009*, pages 91–98, November 2009.
- [3] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proceedings of the 1st conference on Visualization '90, VIS '90*, pages 361–378, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.
- [4] M. John, C. Tominski, and H. Schumann. Visual and analytical extensions for the table lens. In *Proceedings of IS&T/SPIE Annual Symposium Electronic Imaging – Visualization and Data Analysis (VDA '08)*. SPIE Press, 2008.
- [5] R. Kosara, F. Bendix, and H. Hauser. Parallel sets: Interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics*, 12:558–568, July 2006.
- [6] A. Lex, M. Streit, C. Partl, and D. Schmalstieg. Comparative analysis of multidimensional, quantitative data. *IEEE Transactions on Visualization and Computer Graphics*, 16:1027–1035, November 2010.
- [7] R. Rao and S. K. Card. The table lens: merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In *Conference companion on Human factors in computing systems, CHI '94*, pages 318–322, New York, NY, USA, 1994. ACM.
- [8] A. Telea. Combining extended table lens and treemap techniques for visualizing tabular data. In *Proceedings of Joint Eurographics – IEEE VGTC Symposium on Visualization (EuroVis '06)*, pages 1–9, 2006.
- [9] X. Yuan, P. Guo, H. Xiao, H. Zhou, and H. Qu. Scattering points in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 15:1001–1008, November 2009.