Interactive Poster: Two-Tone Pseudo Coloring for Multiple Variables

Mathias John
Institute for Computer Science
University of Rostock
mj069@informatik.uni-rostock.de

Christian Tominski
Institute for Computer Science
University of Rostock
ct@informatik.uni-rostock.de

Heidrun Schumann
Institute for Computer Science
University of Rostock
schumann@informatik.uni-rostock.de

ABSTRACT

Two-tone pseudo coloring (TTPC) is an effective technique to visualize one-dimensional quantitative data. TTPC was originally proposed for only one-dimensional data.

In this work, we present novel applications of TTPC. We combine TTPC with a table lens to facilitate the visualization of *n*-dimensional quantitative data. To increase the applicability of two-tone pseudo coloring for data with high frequencies, we incorporate self-organizing maps (SOM). We also consider extensions that aim at improving the readability of TTPC.

1 INTRODUCTION

Recently, two-tone pseudo coloring (TTPC) was introduced as a technique to facilitate visualization of one-dimensional quantitative data [1]. Whereas classic color-coding (irrespective of whether it is done in a discrete or continuous fashion) maps each data value to one color, TTPC uses two colors to communicate a data value. TTPC has the advantage of providing both overview and detail of the visualized data.

Our interest regards the application of TTPC to visualize multiple variables. We describe a visualization tool that combines the table lens approach [2] and the TTPC approach. A challenge is to cope with data that are changing frequently. We explain how such data can be sorted with self-organizing maps (SOM) [3], to avoid high alternations in the visual representation. To further improve the readability of our two-tone pseudo colored table lens, we address color-related aspects and provide additional views.

2 TWO-TONE COLORING FOR MULTIPLE VARIABLES

To apply two-tone coloring for multivariate data, several questions have to be answered. We focus on the following ones:

- How can multiple variables be arranged on the display?
- How can high frequencies be dealt with?
- How can users be supported in interpreting two-tone pseudo colored visualizations?

Next, we will explain possible answers to these questions.

2.1 ARRANGING MULTIPLE VARIABLES

Although a variety of approaches to visualize multiple variables are known in literature, not all of them are eligible for a combination with TTPC. TTPC requires "*a rectangular area for visualization*" [1]. A visualization approach that fulfills this requirement is the table lens [2]. This technique uses a spreadsheet layout to represent multiple variables. In contrast to common spreadsheets, data values are represented by small horizontal bars. The length of a bar encodes a certain data value. This is particularly useful to provide an overview of the data. To facilitate the precise reading of data values, users can select rows to focus on. For these rows, data values are printed, and hence, users can comprehend them easily.

To combine the table lens approach and the TTPC approach, data values are no longer represented by the lengths of bars, but by bars that use TTPC (see Figure 1). The computation of a two-tone pseudo colored bar is analog to the procedure described in [1]. In a first step, the two colors that are adjacent to a particular data value are chosen from a color scale. These two colors provide a coarse indication of the data value (overview). In a second step, the proportion of how many pixels of a bar are colored in either tone is calculated based on the specific data value. The proportion facilitates a more precise identification of the data value (detail).

A table lens that features TTPC is a promising tool to handle data that are both large with respect to the number of tuples and large with respect to the number of variables. The former case is already addressed by the original table lens. The latter case is a result of applying TTPC, because it facilitates the reading of data values even if the columns of the table lens are narrow.

2.2 USING SOM TO REDUCE HIGH FREQUENCIES

High frequencies are a big problem for the TTPC approach, since they make it difficult to see any developing of values. To avoid frequent alternations in the visual representation, data tuples need to be sorted. As our interest regards *n*-dimensional data, a method is required that can sort multiple variables. We apply self-organizing maps (SOM) as proposed in [4] to solve this problem.

SOM are artificial neuronal networks that cluster unstructured multidimensional data efficiently. Originally, they consist of a two-dimensional network of neurons, typically arranged on a regular lattice. Since our goal is to arrange information objects linearly, rather than to organize them on a two-dimensional grid, we use a one-dimensional SOM, which has proved to provide correct orderings as well [4]. In our SOM, each neuron is initially associated with a single randomly chosen *n*-dimensional data vector. In a second step, the neuronal network is trained with further data vectors (called training vectors) several times. In each iteration, the SOM is searched for the neuron (called the winning neuron) that is most similar to the current training vector. The winning neuron is updated such that it represents the current training vector more closely. Along with that, the neurons in the neighborhood of the winning neuron are adjusted in response to the actual training vector. After the training phase, each neuron represents a certain number of data vectors that are close (i.e., similar) in information space. This enables us to sort data tuples according to their affiliation to the neurons (see Figure 2).

Although SOM deliver good results for the frequency problem, it is possible to improve the sorting. For doing so, the data tuples are additionally sorted within each SOM cluster. We implemented a linear and an agglomerative inter-cluster sorting. The linear method is a straight-forward sorting. The agglomerative method uses a dissimilarity matrix for grouping data tuples hierarchically [5]. Both the linear and the agglomerative method rearrange data tuples according to their similarity within a SOM cluster. Furthermore, the agglomerative algorithm provides a binary tree with clusters as nodes and data tuples as leaves. This tree can be used to drive an interactively controllable semantic zoom of the table lens.

2.3 IMPROVING THE READABILITY OF TWO-TONE COLORING

Besides increasing the applicability of TTPC by sorting the data, it also makes sense to incorporate further means to improve the readability of a two-tone pseudo colored table lens. First of all, the use of color has to be considered. To support the identification of particular variables, we try to represent each variable with a dedicated color scale (see Figure 1). Apparently, this is not possible for datasets with more than about eight variables. Nonetheless, we assure that adjacent columns are represented with color scales that can be easily distinguished. For our implementation, we use evaluated and commonly accepted color scales as provided by the Color Brewer [6].

Secondly, an important aspect is to determine the data values for which colors are switched, so that users can easily interpret the TTPC (e.g., color switches at 10, 20, and 30, rather than at 7.45, 17.45, and 27.45). For one-dimensional data, it is acceptable that this task is performed interactively by users. When considering arbitrary n-dimensional data, simply dividing value ranges into a certain number of segments is not sufficient. We apply a heuristic method to compute the segments. Depending on data characteristics, minimum and maximum of the variables' value ranges are expanded such that the division into segments results in segment borders that are easier to comprehend (see Figure 1).

As already indicated, it makes sense to take advantage of the binary tree that is computed during the agglomerative inter-cluster sorting. For doing so, we provide an additional view that can be used to interactively expand or collapse particular nodes of the binary tree. This allows users to explore the data at different levels of aggregation (see Figure 3).

3 Summary

In this work, we have introduced a combination of the table lens approach and the two-tone pseudo coloring approach. Our prototype software incorporates sorting of data with self-organizing maps to cope with high-frequency data. We also included methods to improve the readability of our two-tone pseudo colored table lens.

In future work, we will incorporate methods to sort data columns according to similarity of variables, and we plan to extend the TTPC approach to categorical data. We hope to find better heuristics to improve the computation of color segments. It is also interesting to combine our method with other table lens-based visualization approaches (e.g., Treemap Table Lens [7]).

References

[1] Saito, T., Miyamura, H.N., Yamamoto, M., Saito, H., Hoshiya, Y., Kaseda, T.: Two-Tone Pseudo Coloring: Compact Visualization for One-Dimensional Data. *Proc. IEEE Symp. on Information Visualization*, Minneapolis, 2005.

[2] Rao, R., Card, S.K.: The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus+Context Visualization for Tabular Information. *Proc. SIGCHI Conference on Human Factors in Computing Systems*, Boston, 1994.

[3] Kohonen,T.: *Self-organizing Maps*. Berlin: Springer, 2001.

[4] Kreuseler, M., Schumann, H.: A Flexible Approach for Visual Data Mining. *IEEE Trans. on Visualization and Computer Graphics*, Vol. 8, No. 1, 2002.

[5] Kaufmann, L., Rousseeuw, P.J.: *Finding Groups in Data – An Introduction to Cluster Analysis*. New York: Wiley-Interscience, 1990.

[6] Brewer, C.A., *Color Brewer*, http://www.colorbrewer.org, 2006.

[7] Telea, A.: Combining Extended Table Lens and Treemap Techniques for Visualizing Tabular Data. *Proc. of EG/IEEE VGTC Symp. on Visualization*, Lisbon, 2006.

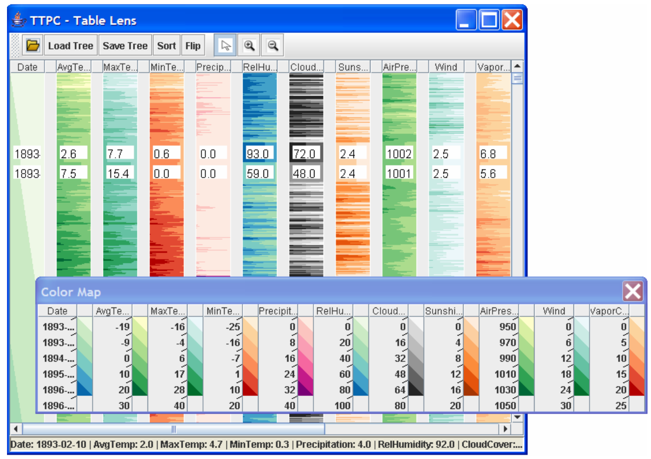

Figure 1 This figure shows our prototype, which combines a table lens with the two-tone pseudo coloring approach. The represented climate data are naturally sorted with respect to time.

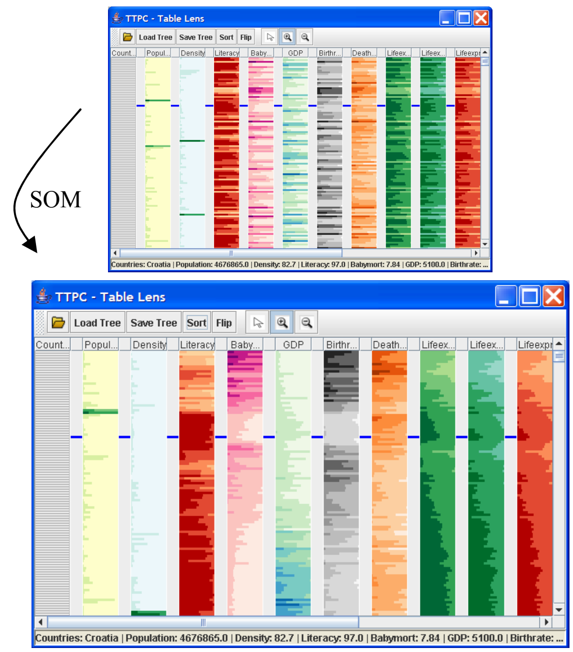

Figure 2 Sorting data tuples with a SOM reduces alternations in the visual representation and helps identifying clusters in the data. The figure shows census data of different states. The SOM sorting reveals a cluster of former East Block states (near the blue line).

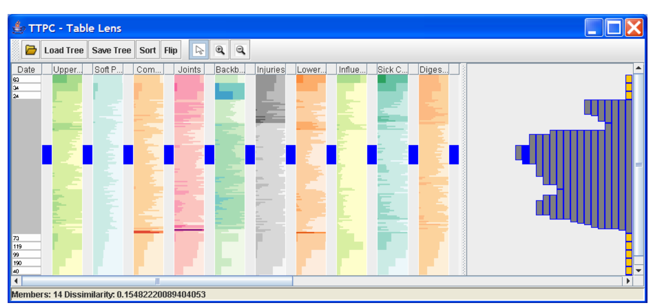

Figure 3 An additional view of the clustering hierarchy (right) can be used to zoom into the data on a semantic level. Larger datasets (e.g., health data as shown here) can be explored more easily.