International Workshop on Visual Analytics (2011)
S. Miksch and G. Santucci (Editors)

Interactive Analysis of Object Group Changes over Time

S. Bremm[1] and T. von Landesberger[1,2] and G. Andrienko[3] and N. Andrienko[3] and T. Schreck[1]

[1] Interactive Graphics Systems Group, Technische Universität Darmstadt, Germany
[2] Fraunhofer IGD, Darmstadt, Germany
[3]Fraunhofer IAIS, Bonn, Germany

Abstract

The analysis of time-dependent data is an important task in various application domains. Often, the analyzed data objects belong to groups. The group memberships may stem from natural arrangements (e.g., animal herds), or may be constructed during analysis (e.g., by clustering). Group membership may change over time. Therefore, one important analytical aspect is to examine these changes (e.g., which herds change members and when). In this paper, we present a technique for visual analysis of group changes over time. We combine their visualization and automatic analysis. Interactive functions allow for tracking the data changes over time on group, set and individual level. We also consider added and removed objects (e.g., newly born or died animals). For large time series, automatic data analysis selects interesting time points and group changes for detailed examination. We apply our approach on the VAST 2008 challenge data set revealing new insights.

Categories and Subject Descriptors (according to ACM CCS): H.5.0 [Computer Graphics]: Information Systems—Information Interfaces and Presentation General

1. Introduction

In this paper, we address the analysis of time-dependent data with grouped objects, which is relevant to various application domains. The groupings may be pre-defined (e.g., animal herds in biology), or may be a result of previous analysis (e.g., clustering). When the group membership changes over time, it is necessary to examine these aspects (e.g., which herds change members and when).

The analysis tasks include:

- tracking group sizes and contents over time,
- tracking (sub)sets of objects (e.g., a set of objects changing the same groups between two point in time),
- tracking group memberships of individuals,
- analyzing group changes (how many group members switch to which group),
- examining objects added to and removed from groups (e.g, newly born or died animals).

The main challenge is the scalability w.r.t. the number of group changes and the number of time points. Therefore visualization combined with data analysis is needed. The related works either focus only on interactive visualization of group changes [KBH06, ZKG09, LSP*10], or only on time development of the group sizes [HHN02].

In our work, we combine data visualization with automatic analysis. The data display extends the Parallel Sets concept [KBH06] with the time dimension and includes additional glyphs for easier identification on group changes. We introduce a) tracking of group changes on three levels: whole groups, sets of objects and individuals; and b) automatic data analysis selecting representative points in time for deeper examination.

We apply our approach on the VAST Challenge 2008 data set [GPL*08]. The new approach allows us to track movements of people across the building areas on the group level. It shows the common and extraordinary movements over various time intervals.

2. Related Work

Our work relates to several areas.
First, a survey of techniques for visual analysis of time-dependent data is presented by Aigner et al. [AMM*08]. Relevant for our topic is ThemeRiver [HHN02] showing developments of theme changes over time. It visualizes the group sizes in a stacked chart. The increasing and decreasing number of objects in each category can be fol-

lowed. History flow [VWD04] shows similarly the development of documents by authors. Thereby, the important themes/authorships in each time point can be identified. They, however, do not include the changes of group membership between time points.
Second, interactive visualization of group memberships in several categories has been presented in [KBH06]. The so-called Parallel Sets show changes of group memberships across several categorizations. Interactive Sankey diagrams [Fro05] address the same issue in a similar way as Parallel Sets, focusing on visual design. Both works do not regard time-dependent data. We build upon this approach extending it with time dimension and further visual and interactive features for better analysis of group changes.
Third, Zhou et al. [ZKG09] and Lex et al. [LSP*10] proposed approaches for partition comparison. They both consider several groupings of data objects according to different clustering results. They connect individual objects across clusterings while minimizing edge crossings. They however do not regard time dependency of the data.

3. Approach

In our approach, we extend Parallel Sets concept [KBH06] with object tracking and automatic data analysis functions for time-dependent data. We assume that the time-dependent data grouping information is given.

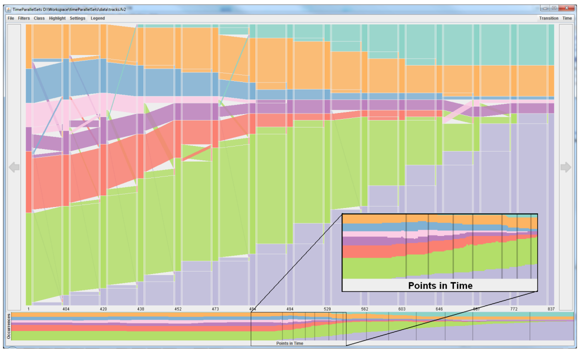


Figure 1: *The main view. Upper part: Group distributions (bars) and group changes (edges) in selected time points. Lower part: Object distribution in the whole time period. Challenge data set showing the changes between 15 automatically selected time points (highlighted by vertical lines).*

The visualization shows two levels of detail (see Fig. 1). The groups are color-coded and shown in user-defined order on vertical axis. On the bottom, the distribution of all objects among groups in every moment is shown. The upper part displays the object distribution (bars) and the group changes of objects (edges between bars) in selected time moments, whose temporal positions are shown in the lower part by vertical lines. A good selection of these points in time is crucial for an effective analysis. Therefore, we included also an automatic data analysis to extracting time points well representing the input data.

3.1. Visual Attributes

In the main view, each vertical bar represents the group distribution in one moment. Edges between neighboring bars represent objects group changes between two points in time. Their color is optionally based on source or destination group. The changes can additionally be shown by the color coding of the narrower bars aside of every group. It indicates the proportional group membership of the contained objects in the time steps before and after the current one (Fig. 2 left). It is especially useful when the outgoing/incoming edges are strongly overlapping. The analyzed dataset does not necessarily contain the same / all objects in every point in time. For representing this fact, added and removed objects are indicated as triangles at the side of the group bars (Fig. 2 left).

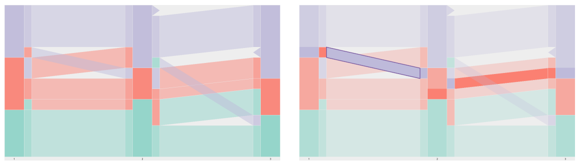


Figure 2: *Example visualization and tracking. Left: Bars show group structures in selected moments. Edges and the side-boxes show sets of objects switching groups. Incoming or exiting objects are indicated by triangles at the sides of the bars. Right: Tracking of sets of objects (switching from blue to red in the first time step).*

3.2. Tracking

To enhance the data analysis, not only the overview over group changes but alsdo the tracking of sets of interesting objects over time is important. Tracking encompasses three different aspects of the data which are highlighted:
1. *Tracking of groups:* The bars of the selected groups are highlighted as well as all leaving edges. The user can comprehend the characteristics of this group and the contained objects, e.g., if leaving objects always swap together to one group or not.
2. *Tracking of individuals and object sets:* An object or a set of objects can be selected, e.g., by clicking on an edge. Consequently, all bars, side bars and edges containing the selected objects are automatically highlighted proportionally to group size. In this way, the transitions of a set of objects across groups can by examined, e.g., to check if the set objects have common transitions (Fig. 2 right).

3.3. Automatic Highlighting and Filtering

A good selection of the points in time for a detailed analysis is important. It should represent the data well. We support automatic selection of interesting points in time (Fig. 3):
1. As a starting point for the analysis, an *automatically generated overview* of the data is provided. Instead of equally sized time intervals, the time selection equalizes the total number of objects changing groups in each selected time interval.Thus, overall trends are visible and time periods of high activity (shorter intervals) are identified.
2. An additional *outliers search* provides a set of moments with extraordinary group-change events. Moments are selected, when the number of objects changing groups is below a user-defined threshold (i.e., seldom changes). The user can recognize events which otherwise could not be seen in the time series overview.
3. The *transition matrix* shows the number of group changes aggregated over time as a heat map. The cells indicate common and extraordinary group changes. They are linked to the main view for highlighting the selected transitions.

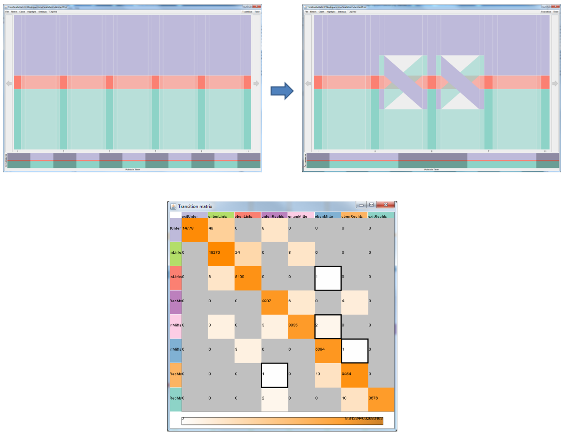


Figure 3: *Automatic data analysis. Top: Identification of interesting time points based on group changes. Equal time distribution vs. relevance-based time selection. Bottom: Transition matrix showing the amount of group changes.*

4. Application

We demonstrate the usefulness of our approach on “IEEE VAST 2008 Challenge – mini challenge 4: evacuation data” [GPL*08]. The dataset describes the movement of 82 subjects in a building over 837 points in time [AA10]). We partitioned the building into eight areas (Fig. 4 left) and put the subjects into groups according to their location in every moment. The group coloring is seen in Fig. 4 right. At a specific time, a bomb detonated and afterwards people die or start to move towards the exits (turquoise and purple areas).

The input data is visualized on the basis of 15 automatically recommended time points, which are highlighted in the overview on the bottom (Fig. 5(a)). The time distribution is unequal over the whole time period. The automatic analysis puts more emphasis on time periods of high movements (after the explosion). As can be seen, most people reached one of the two exits. To reveal the escape routes, the subjects arriving in the particular exits are highlighted. People using the bottom exits were at the beginning mostly located in the upper left or one of the bottom regions. Many people move as expected towards the nearest exit, e.g., from the top left (red) to the bottom left (green) region. Surprisingly, even some people from the top middle region (blue) preferred the bottom exit to the closer right exit (Fig. 5(b)).

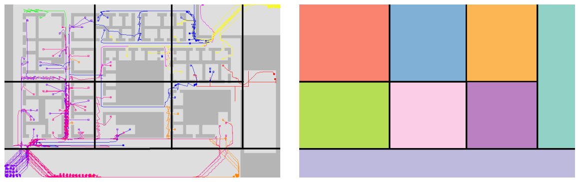


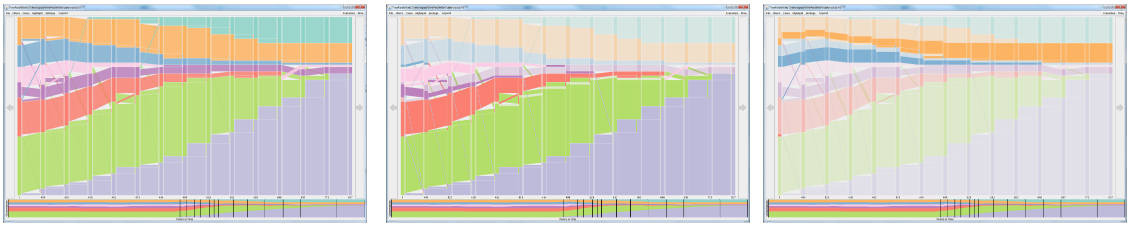
Figure 4: *The example dataset from the VAST challenge 2008 [GPL*08]. (a) Trajectory based visualization of the movement of the people [AA10] with partitioning of the building into 8 areas. (b) Coloring of the areas.*

If a security analyst wants to improve the escape routs, she can first highlight all people, who remain in the upper right (orange) area after the evacuation (Fig. 5(c)). Some of them have never moved (probable victims of the bomb). Interestingly, there is one person who came from the upper mid (red) area. He did not take the shortest path so he did not manage to exit the building fast enough.

She can then focus on the analysis of the behavior of people who move differently from the rest or in an unexpected way (away from exits). Those few region changes are selected in the transition matrix (Fig. 3 bottom) and are then highlighted in the main view. As it turns out, only some of the marked transitions (orange to blue and blue to red) are visible (Fig. 6(a)) in the initial view (using automatic time selection). The other transitions (purple to orange and blue to pink) are not captured by the selected points in time. Therefore, the selection of the representative time points was recalculated, ensuring to show all selected transitions and at the same time achieving a overview about the whole data with the desired 15 key points in total (Fig. 6(b)). Tracking these people reveals that despite their odd routing, the majority reaches the exits (Fig. 6(c)).

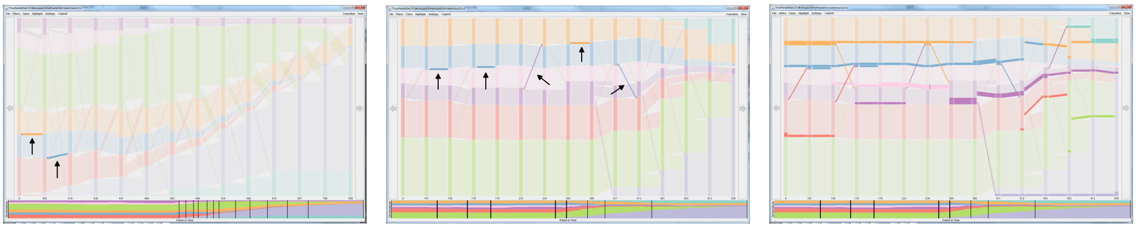
5. Conclusions and Future Work

We have presented an approach for visual analysis of time-dependent data with non-constant grouping. The approach



(a) (b) (c)

Figure 5: *Visualization of the evacuation data. (a) Data visualization using relevance-based time selection. (b) Tracking of people who reach the exit on the bottom. (c) Tracking of people who were located in orange area at the end.*



(a) (b) (c)

Figure 6: *Analysis of unexpected transitions. (a) Initial time selection shows only some of the selected transitions. (b) Overview of interesting transitions using transition-based time selection. (c) Tracking of selected transitions. Despite the unexpected movement, most of the highlighted people reach the exit in time.*

allows for tracking of group changes over time. It combines interactive visualization with automatic data analysis for revealing interesting insights into data.

In the future, we would like to extend our system with automatic group ordering and pattern recognition as well as evaluation of the scalability w.r.t. number of groups and objects.

Acknowledgements

This work was partially supported by DFG SPP 1335 projects "Visual Feature Space Analysis" and "ViAMoD" as well as the German BMWi THESEUS program. We thank Kerem Gülensoy and Matthias Götz for their help with the project.

References

[AA10] ANDRIENKO G., ANDRIENKO N.: Interactive cluster analysis of diverse types of spatio-temporal data. *SIGKDD Explor. Newsl. 11* (May 2010), 19–28.

[AMM*08] AIGNER W., MIKSCH S., MULLER W., SCHUMANN H., TOMINSKI C.: Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics 14*, 1 (2008), 47 –60.

[Fro05] FROEHLICH P.: Interactive Sankey Diagrams. In *IEEE Symp. on Information Visualization* (2005), p. 233.

[GPL*08] GRINSTEIN G., PLAISANT C., LASKOWSKI S., O'CONNELL T., SCHOLTZ J., WHITING M.: VAST 2008 Challenge: Introducing mini-challenges. In *IEEE Symp. on Visual Analytics Science and Technology* (2008), pp. 195–196.

[HHN02] HAVRE S., HETZLER B., NOWELL L.: ThemeRiver: Visualizing theme changes over time. In *IEEE Symp. on Information Visualization* (2002), pp. 115–123.

[KBH06] KOSARA R., BENDIX F., HAUSER H.: Parallel sets: Interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics* (2006), 558–568.

[LSP*10] LEX A., STREIT M., PARTL C., KASHOFER K., SCHMALSTIEG D.: Comparative analysis of multidimensional, quantitative data. *IEEE Trans. on Visualization and Computer Graphics 16*, 6 (11 2010), 1027–1035.

[VWD04] VIEGAS F. B., WATTENBERG M., DAVE K.: Studying cooperation and conflict between authors studying cooperation and conflict between authors. In *CHI* (2004), pp. 575–582.

[ZKG09] ZHOU J., KONECNI S., GRINSTEIN G.: Visually comparing multiple partitions of data with applications to clustering. In *Proceedings of SPIE* (2009), vol. 7243, p. 72430J.