

Using Visual Analytics to Provide Situation Awareness for Movement and Communication Data

Juri Buchmüller*

Fabian Fischer†

Dirk Streeb‡

Daniel A. Keim§

Data Analysis and Visualization Group
University of Konstanz, Germany

ABSTRACT

Analyzing and correlating movement and communication data is challenging. To gather insights and gain knowledge out of such datasets, we propose a visual analytics system. We apply automated clustering techniques and propose a combination of various visualizations to provide overviews. We support the analyst in exploring the data to eventually enhance situational awareness in complex analysis scenarios. To evaluate our approach, we apply our techniques in the context of the VAST 2015 Grand Challenge (GC). Within this challenge, we successfully identify suspicious patterns and interesting groups of distinctive behavior among visitors of an amusement park, and correlate them with their respective communication patterns to gain insights.

Index Terms: H.2.8 [Database Management]: Database Applications—Data mining; H.5.2 [Information Interfaces]: User Interfaces—Graphical user interfaces (GUI)

1 INTRODUCTION

This paper accompanies our submission for the VAST 2015 Grand Challenge (GC). For joint inspection of movement and communication data, we combined a collection of analysis and visualization tools. The available dataset comprised three days of movement and communication data among visitors of a fictional amusement park. This data relates to an app used by most of the visitors to send text messages to other visitors.

2 BACKGROUND

Our approach is designed after the principles of visual analytics as defined by Keim et al. [2], and its workflow follows the knowledge generation model [3] proposed by Sacha et al. [3], who describes the fundamental interaction between human and system. The analyst goes through loops of exploration and verification leading to insights, findings, and eventually generates knowledge. On the system side, we mainly focus on the three areas of data handling, model generation, and novel visualizations. For the first two we build upon KNIME [1], a flexible data mining tool. Our prototypes aim to fill the gap on the visualization side. In combination, we present a powerful visual analytics system for the analysis of movement and communication data to identify patterns and groups of similar behavior.

3 VISUALIZATIONS

Our visual analytics system consists of different main visualizations including global pattern heatmaps, communication network

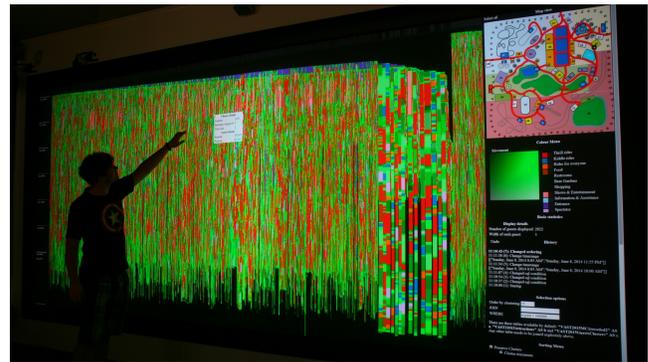


Figure 1: **Visual Movement Explorer** – A visual analytics system showing large amounts of movement data to visually explore group behavior. Each colored column represents a visitor over time. Movement is represented as shades of green, other colors belong to activities within the park (e.g., attended attractions). Interaction, sorting, filtering, and clustering can be used to explore the data.

graphs, and the *Visual Movement Explorer* as shown in Figure 1. While it is possible to use the visualizations individually, it is the combination that makes it possible for the user to verify findings and gain insights from multiple perspectives.

Global Pattern Heatmaps – To get a first overview, we implemented several heatmap-based visualizations, with each of them focusing on one type of data, including attraction check-ins, guest movements and communication activity. Each of the heatmaps provides visual access to global phenomena or can be used to compare various sub-groups to each other with respect to the global pattern. The overall attraction check-in heatmap shows the occupancy rate of attractions, and eventually correlates to their opening hours. When comparing subgroups, individual interests, preferences and dislikes in specific types attractions become visible.

Communication Networks – To analyze communication data, we included a node-link diagram to present the network graph, and a specialized visualization showing message flow along parallel coordinates (Figure 2). While the network graph tells who communicates with whom, the message flow visualization shows the temporal component and whether a conversation is uni- or bi-directional. This way, one-shot messages can be distinguished from proper conversations. Additionally, the whole conversation flow from receiving a message and answering the same person to spreading the message to further people becomes visible. Together, these visualizations enable the analyst to get an overview and inspect interesting communication patterns in more detail.

Visual Movement Explorer – To analyze movement in detail, we implemented a custom visualization, called the *Visual Movement Explorer*, aiding in the analysis and profiling of large amounts

*e-mail: buchmueller@dbvis.inf.uni-konstanz.de

†e-mail: fischer@dbvis.inf.uni-konstanz.de

‡e-mail: dirk.streeb@uni-konstanz.de

§e-mail: keim@dbvis.inf.uni-konstanz.de

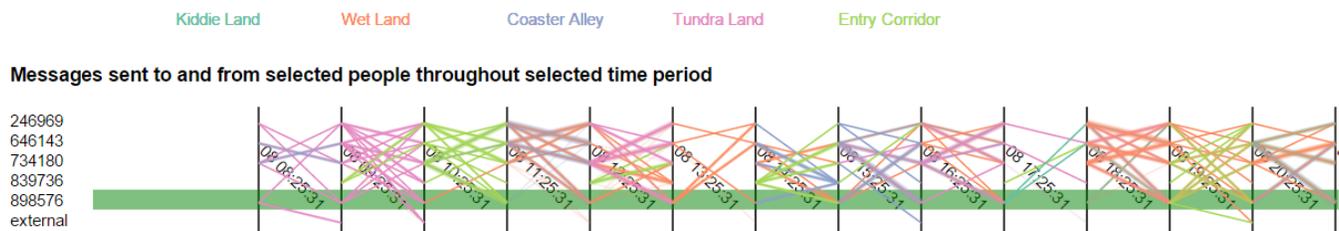


Figure 2: **Communication Flow Visualization** – This visualization can be used to explore the communication of multiple visitors in the amusement park. The parallel coordinates based representation shows communication between people over time. The scale on each vertical axis always relates to the list of 5 visitors as seen on the left side. Each vertical axis represents a specific time span. The number of messages is mapped to the width of lines, while color relates to geographic region from where a message is sent.

of movements in one screen by providing a context as complete as possible (Figure 1). The visualization shows each visitor as a colored column. Time is mapped to the vertical axis. So, beginning at the top and going down a column to the bottom one can follow a guest through a selected timeframe. A marker links the actual position on a map on the right hand side. Where a guest is and what he does is coded to color. In Figure 1 the different shades of green denote movement in various areas of the park. Other colors reveal how much time the visitor spends within the various attraction types. Interactive selection and highlighting can be used to ease recognition of patterns. The order of the different columns (visitors) can be steered interactively by the user. In the picture you can see a simple ordering by the time guests enter the area respecting clusters. All extracted features can be used to sort and also cluster the view, to provide means to visual explore the behavior groups. Following the basic idea of interactive exploration almost any preset can be altered on the fly.

4 WORKFLOW TO IDENTIFY SUSPICIOUS VISITORS

One task within the VAST 2015 Grand Challenge was to find out when and where crime happened and to suggest some suspects. Looking at the global heatmap of attraction check-ins, we can spot different usage patterns across days. One interesting change is related to the *Creighton Pavilion*, because it seems to be closed on Sunday afternoon.

With this as a first finding we now take a look at the global communication heatmap. There we can find a peak beginning some time before the pavilion is closed. Taking a look at the emerging peak we can see that it actually started in the area of the park where the pavilion is located.

Thus taking a closer look at what people do at that time appears to be appropriate. Thinking about what a criminal probably would do, we can use the *Visual Movement Explorer* to identify visitors leaving the park around that time. The visualization allows us to compare their movement not only between each other, but also with previous days.

Looking at the communication data, socially-connected network groups can be identified within those leaving the park. Interestingly, some of them do not communicate at all. Visitors with strange or no communication patterns might be more suspicious than others. But who of them was actually able to commit the crime?

By selecting an area around the crime scene, we can easily see who was close to the pavilion during the respective time span. Some of them show very common behavior patterns, and others are too far away, so we can justifiably exclude them from our list of suspects. One visitor, however, seems to spend all the time at a barbecue stall nearby the pavilion. He could have left his tracking device there and have picked it up after having committed the crime. Though if he communicated in the time in question we could reject this hypothesis. Our communication visualization shows that this person does not communicate in the given time — actually not a single

time at all. Looking at his behavior once again we additionally see that it was similar on Friday and Saturday yet changed on Sunday.

Finally with the help of our visual analytics system, we found various highly suspicious persons, that could be inspected further, ranked, and handed to police officials.

5 CONCLUSION

Our visual analytics system provides specialized, interactive visualizations specifically designed to identify group behavior and individuals that do not fit common patterns, both in movement and communication. The system supports the analyst in building hypotheses using automated clustering and interactive visualizations to verify insights and findings.

Secondly, the variety of different visualization perspectives helps to focus on different tasks and questions from the same starting point, enabling the user to go into detail whenever needed, to verify insights and to validate hypothesis.

Using the challenge dataset, we were able to address the grand challenge with very promising results. Furthermore, we strongly believe that the visualizations are general enough to be applied to other scenarios with similar datasets.

ACKNOWLEDGEMENTS

We would like to thank the additional students who were working hard and motivated on individual contributions for the specific mini challenges, namely Carmela Acevedo, Eren Cakmak, Alexander Gärtner, Thomas Hepp, Udo Schlegel, and Bruno Schneider.

REFERENCES

- [1] M. R. Berthold, N. Cebon, F. Dill, T. R. Gabriel, T. Kötter, T. Meinel, P. Ohl, C. Sieb, K. Thiel, and B. Wiswedel. KNIME: The Konstanz Information Miner. In *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*. Springer, 2007.
- [2] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. *Visual analytics: Definition, process, and challenges*. Springer, 2008.
- [3] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge Generation Model for Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visual Analytics Science and Technology 2014)*, 20(12):1604 – 1613, Dec. 2014.