

Using Visual Analytics to Analyze Movement and Action Patterns

Dirk Streeb

Udo Schlegel

Juri Buchmüller

Fabian Fischer

Daniel A. Keim

Data Analysis and Visualization Group*
University of Konstanz, Germany

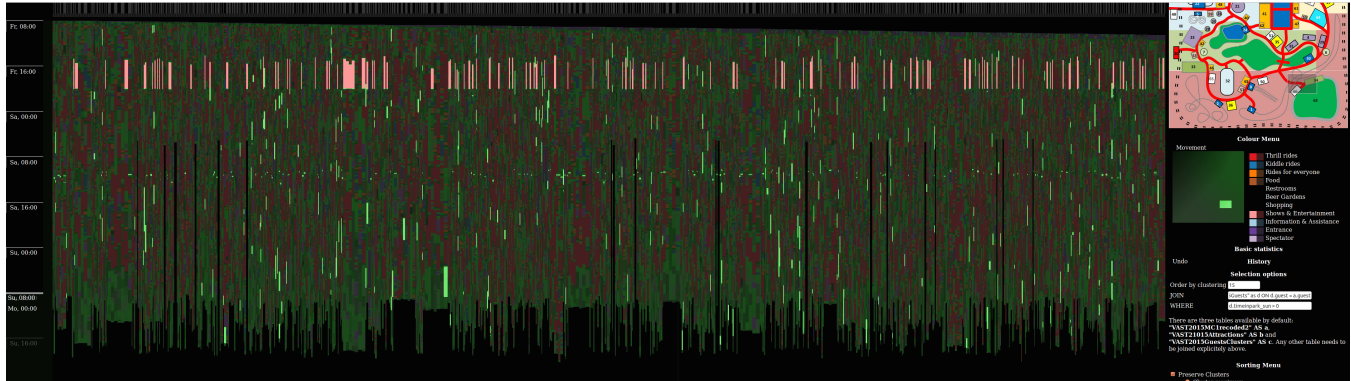


Figure 1: **Visual Movement Explorer**: The columns on the horizontal axis represent visitors. The colored segments on the vertical axis in each visitor column relate to the type of activity for the respective visitor. Areas on the map can be selected and all other colors not corresponding to the area get darker. The view can be sorted and grouped according to clustering results based on extracted features to differentiate between groups with similar behavior.

ABSTRACT

This paper describes our approach and tools used to solve this year’s VAST 2015 Mini-Challenge 1 (MC1). It also explains how we came up with the idea for a new tool created for solving of the challenge and how this tool helps the user analyzing the data. Further, preprocessing steps will be discussed and how new features were created to be able to cluster guests in groups walking together. Finally findings of this visualization will be presented.

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 completes our submission for the VAST 2015 Mini-Challenge 1 (MC1). This year’s challenge “Mayhem at DinoFun World” was about crime and mayhem in a fictitious amusement park. While a famous soccer player was holding a show for the visitors, a pavilion was vandalized and trophies and awards of the soccer player were stolen. Mini-Challenge 1 involves the analysis of individual and group movement against a dynamic background environment. This paper is a solution to solve Mini Challenge 1 using a visual exploration tool based on the visual analytics approach.

To solve the given task, the data mining and analysis tool KNIME [1] will be used to preprocess the data, which lays the foundation for a novel visualization technique. This technique, called the Visual Movement Explorer, will be explained and presented below, together with the related knowledge discovery workflow it is associated with.

*e-mail: firstname.lastname@uni-konstanz.de

2 DATA ANALYSIS WITH KNIME

Movement data for Mini-Challenge 1 contains only guest ids, coordinates in the park area, timestamps and check-in markers for attractions. So, it is not obvious who moves with whom in the first place. To use our visualization it is however necessary to define these groups beforehand using a clustering approach. Prior to that, we derive the necessary features from the data using KNIME.

We decided to divide these new features in movement based features, time based features, and attraction based features. For movements, a central feature is speed. It was calculated for each guest to get an impression if they are walking fast or slow and especially to find guests which have similar movement speed. Further, walked distance was calculated for the three days and every guest, as groups are most likely to walk a similar distance at least within a day. In some cases groups were found with some group members staying an additional day while others left the day before. For time features, time in park was calculated in a similar way as walked distance. Also, time difference between the first check-in in an attraction and entering the park was calculated for all three days, as groups should have a similar time difference there. The amount of days in the park was also calculated to easily separate guests who stay in the park longer on one day from those staying a shorter time on each of two days. With regard to attractions the time spent in an attraction was calculated. This includes the time waiting to get in as well as riding itself. Furthermore, it was calculated how often guests checked in at different attractions to derive groups that for example like to ride thrill rides. All these features were calculated as precise as possible to maximize inter-group variance. As we look for groups spending most of their time together we expect intra-group variance to be less affected.

The DBSCAN [2] clustering algorithm was used with various parameters combinations for clustering. Cluster quality was evaluated visually using visualization. Our final combination came up with only around 2200 guests as noise or people without being in a

group and around 2200 groups including about 10000 guests. Some further refinements could have improved the clustering further, yet expected gains did not exceed the time needed for implementation.

3 VISUAL MOVEMENT EXPLORER

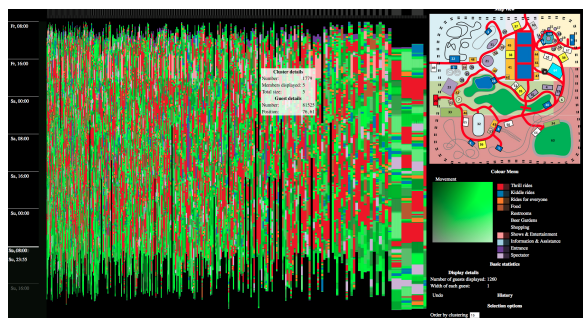


Figure 2: Visual Movement Explorer showing guests on Sunday, who are in groups in the park. On the right side there are larger groups.

To validate our clusters and to get an idea how the people walked through the park, we introduce a novel visualization, the Visual Movement Explorer (Fig. 2). Each attraction category – except restrooms, beer gardens and shopping attractions – are assigned to a unique color. The park map is mapped to a two dimensional color gradient. This scale uses HSL color space with low lightness in the left top corner, high lightness in the right bottom corner, high saturation in the right top corner and low saturation in the left bottom corner. Despite the drawbacks of the HSL color space we chose it, as linearity is not essential and gradient generation based on a selected hue is intuitive and it shows intuitively areas with a difference in movement for some guests.

To focus on specific aspects any selection of colors and areas on the map can be highlighted. All other events that are not in selected classes or area get displayed darker and less saturated. Interacting via several steps of highlighting and selecting is key in using our visualization.

On the left is a time scale showing the total time span for data is available. Through this scale the time frames to be shown in the main area can be selected and the inspected time can be interactively changed during the course of exploration. On the scale next to it the dimensions of these selected frames is shown, as these are scaled to fit available screen size. By hovering over the area right next to the time scale a line can be seen which goes horizontal over the guests and counts were the guests are on this specific time. The results are shown in the map on the right as a heatmap by darkening the map and drawing a heatmap over the darkened map.

Guests are visualized in the middle. Each column visualizes the movement and check-ins of one guest. The colors are according to the colors of the legend explained above. By hovering over a point in the column, a tooltip is shown with the guest id, the cluster number, the cluster size and the attraction or movement point the guest is currently visiting. Also, the position is shown on the map on the right when hovering, so it is possible to see how the guest moved through the whole park. Above the guests, the clustering is shown in a simple two layer dendrogram. The difference in gray and black shows cluster assignment for each guest.

Guests and clusters can be sorted and selected via a flexible interface offering direct database access and predefined sorting functions. The advantage of the Visual Movement Explorer is that guests with extraordinary movement can be seen and discovered easily. Also, unusual patterns like a closing of an attraction type or a higher usage of an attraction type can be discovered. Even more fine grained attraction selection is possible conceptually, but

not implemented. General overall movement patterns at a specific time can be spotted and further analyzed.

4 FINDINGS

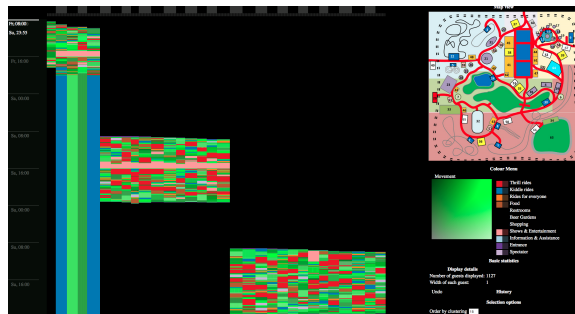


Figure 3: Visual Movement Explorer showing the so called bus groups on all three days, which heavily exceed the second highest cluster sizes.

Through the features, which were calculated and added to the movement features, the Visual Movement Explorer can be used to explore different aspects. For example, by selecting only the movement speed of the guests and circling through different groups to see notable differences. Higher movement speed leads to more thrill ride visits through all guests. Guests with slower movement speed are more likely to be families as they prefer kiddie rides and rides for everyone.

Further, it is possible to show guests via the cluster sizes. So, if you want to find groups with a greater number of guests, you just search for the biggest clusters. In the data there are a few big groups (Fig. 3), who were called bus groups in this solution, because they move exactly the same for their whole stay in the park and only visit one show at the afternoon. These groups were found through looking for group sizes larger than 20, as the second largest type is usually around eleven guests. Outliers were found through looking for people who had special events in the movement, like staying a while in a specific area or attraction or if they have no further movement data later than a timestamp. However, outliers can also be found by selecting attractions with a lower check-in number and look for people, who have a high check-in number there. The body-guard group was found this way, as they attend 5 shows on 3 days, which is not matched by other guests.

5 CONCLUSION

To conclude, a good clustering is necessary for finding groups, but to get interesting results a visual representation is important, as it supports finding groups and outliers by visually exploring the data. For this task our approach of the Visual Movement Explorer was created and successfully used to find groups and patterns in the data. Also, hypotheses which were brought up in the beginning of the challenge could be explained and proofed and others could be discussed and discarded with the Visual Movement Explorer.

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] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *KDD*, volume 96, pages 226–231, 1996.