Spectrum: A Visual Analytics Tool to Explore Movement Logs

Junpeng Wang[*]   Ji Wang[†]   Chris North[‡]

Department of Computer Science, Virginia Tech

ABSTRACT

We present a visual analytics tool, called *Spectrum*, to analyze the movement and communication log data from VAST Challenge 2015. *Spectrum* has two views: *MoveView* and *SpectrumView*. *MoveView* gives an overview of the movement logs at a certain timestamp by synthesizing time, location and identity information. It replays movement logs over time and demonstrates communication logs with dynamic links. *SpectrumView* shows the status of all visitors' activities within a period of time. Each stay of visitors in a location is visualized as a line segment. The duration and location type of the stay are encoded by the length and color of the segment. Accumulating segments of the same visitor horizontally and aligning sequences of segments from different visitors form a general view of all visitors' activities. In our attempt to solve the VAST Challenge 2015, we find that *Spectrum* provides an intuitive overview of the log data and helps to discover movement patterns from a large number of log records.

1 INTRODUCTION AND PROBLEM OVERVIEW

The VAST Challenge 2015 describes a fictitious scenario in which an incident occurred in a typical modest-sized amusement park during a three-day weekend in June 2014. The incident marred the celebration of a famous soccer star who was scheduled to give two stage shows each day. The challenge dataset provides movement and communication logs of all visitors in the park. The movement logs record the visitors' locations and activities in different timestamps; whereas the communication logs record when, where and who communicated in the park. The goal of this challenge is to leverage visual analytics approaches to discover movement patterns of visitors from the large number of log records. To solve the challenge, we propose a visual analytics tool, called *Spectrum*, which replays visitors' activities in the park and clusters visitors according to their movement and communication patterns. We will introduce two views, *MoveView* and *SpectrumView*, of our tool and how they can help to find movement patterns in detail in this paper.

2 MOVEMENT VIEW

MoveView provides an intuitive overview of the movement and communication logs from the VAST Challenge 2015. We introduce three principal features of this view with several examples in this section.

2.1 Movement Replay

Each record in the movement data has four attributes: "time", "id", "event type" and "location coordinates" (e.g. "2014-6-06 08:00:16","1591741","check-in","63,99"). In order to present a snapshot of the park at a specific timestamp, *MoveView* extracts the records of that timestamp and pins them as black dots onto the park map according to their location coordinates. The popularity

---

[*]e-mail: junpeng@vt.edu
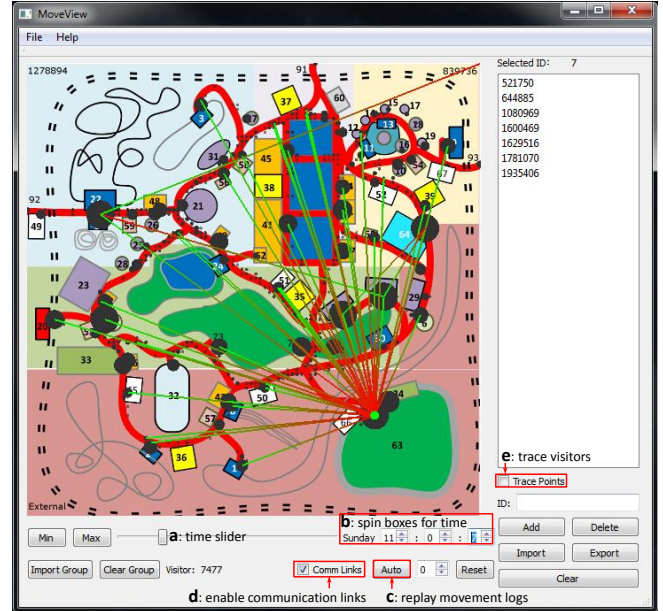
[†]e-mail: wji@vt.edu

[‡]e-mail: north@vt.edu

Figure 1: *MoveView*: the overview of the park at 11:00:07 AM.

of different locations is indicated by the size of the dots. For example, location 63 attracted a large number of visitors at 11:00:07 on Sunday, as shown in Figure 1. The movement logs are discrete records. A record of a visitor was generated only when the visitor's location or event type (*movement* or *check-in*) changed. As a result, some visitors' location information may be recorded in previous timestamps. For example, if a visitor has two consecutive records showing he/she was in location *A* at 10:00:00 and in location *B* at 10:05:00, we will assume the visitor stayed at *A* from 10:00:00 to 10:04:59 and moved to *B* at 10:05:00.

MoveView replays movement logs and provides an intuitive overview of visitors' activities in the park over time. Users can control the replay by: (1) dragging the time slider (Figure 1 *a*); (2) precisely changing the time using the spin boxes (Figure 1 *b*); or (3) enabling automatic replay (Figure 1 *c*). This overview reveals many important activity patterns in the park. For example, around 11:00:00, many dots started flowing out from location 63. We could infer that the stage show just finished around that time.

2.2 Interactive Visual Analysis

MoveView provides multiple interactions to analyze the activities of visitors in the park. In *MoveView*, we can select and track a particular group of visitors to explore their movement patterns. For example, in Figure 1, seven visitors are selected and highlighted in green. They were in location 63 at 11:00:07 on Sunday. By tracking their movements, we can figure out the route they followed from the park gate to this location. We can also export these IDs and track their movements on other days. In the scenario of the challenge, a group of people worked with the soccer star who gave two stage shows in location 63 during specific time periods of each

day. We find this group by: (1) replaying the movement logs to detect the start and end time of these two shows; (2) selecting all visitors who stayed in location 63 during the show time across three days and exporting them into separated sets (one set per show); (3) finding the visitors who appeared in all these sets; (4) tracking the movements of the visitors from (3) to verify the preliminary result.

2.3 Dynamic Communication Links

We find three IDs, i.e. *1278894*, *839736* and *External*, that only appear in the communication logs. Since these IDs are not in the movement data, they have no location information. We place them in three corners of the map (Figure 1). Visitors will be linked on the map if they have communications at the specified timestamp. The direction of a communication link is encoded by colors (from red to green). There might be many communications between two locations. This amount will be denoted by the width of the link. For example, in Figure 1, we can find that a large amount of messages were sent out from location 63 at 11:00:07. These communication links helped to discover the broadcasting communication pattern of ID *1278894*.

3 SPECTRUM VIEW

Different from *MoveView* which presents the geographical status of park visitors at a timestamp, *SpectrumView* gives an overview of all visitors' activities in a time range. In the range, visitors accessed a sequence of locations and stayed varying durations in these locations. We use a line segment to represent a stay. The duration of the stay is encoded by the length and the category of the location is encoded by the color of the line segment. Connecting all segments of a visitor horizontally in time order demonstrates the visitor's activities in the time period. Figure 2 illustrates how we get the sequence of segments from a visitor who visited location 93, 15, 81, 63, etc. Stay sequences from different visitors are aligned vertically in *SpectrumView*.
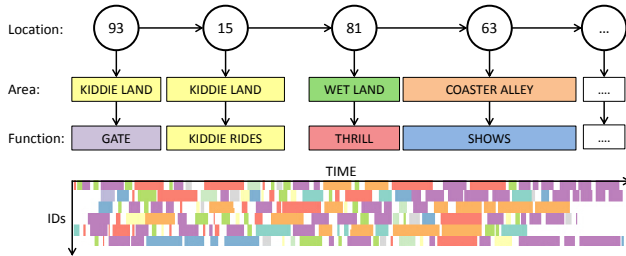

Figure 2: *SpectrumView*: visitors' stay sequences in a time range.

3.1 Filtering

SpectrumView allows users to filter locations based on the category of their *Area* or *Function* (Figure 3 *a*). A stay in *SpectrumView* can be colored either by: (1) the *Area* that the corresponding location belongs to (e.g. Wet Land); (2) or the *Function* of the corresponding location (e.g. Shopping). For example, visitors in the black box of Figure 3 spent most of their time in *Tundra Land* and *Wet Land* locations on Sunday and were very interested in *Thrill Rides*. One or several types of locations can be disabled from the rendering, which helps to discover potential patterns from visual clutter. For example, visualizing *Kiddie Rides* locations and disabling all other types of locations reveal some potential groups with children.

3.2 Sorting and Clustering

Different visitors' stay sequences have different features, like the duration of each stay, the number of visited locations, the type of
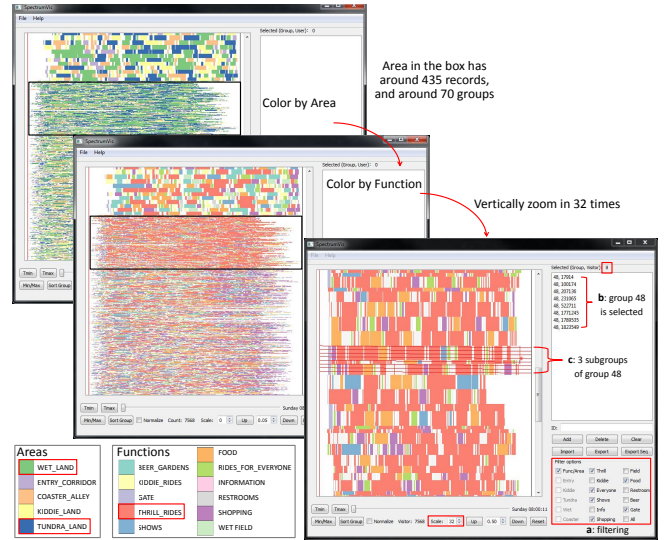

Figure 3: Cluster visitors and filter stays in *SpectrumView*.

visited locations, etc. *SpectrumView* can sort and cluster visitors according to these features. For instance, visitors can be sorted by the time they entered into the park. Since visitors of the same group usually move together, this sorting will put these visitors close to each other in the vertical dimension. Also, the size of the group can be easily identified in the visualization. Sorting by the longest stay of visitors' sequence can expose visitors who stayed in a location for unusual long time. These visitors were more likely to be relevant to the incident. Many existing clustering algorithms, like community detection algorithms [1] and *k*-means, can be used to cluster visitors according to the features of their stay sequence.

In Figure 3, visitors are grouped together if they have communication links. Inside each group, visitors are sorted by the start time of each stay in their stay sequence. The visualization results indicate that visitors in the same communication group may not always move together. The eight selected visitors in Figure 3 *b* separated into three subgroups (Figure 3 *c*) after they entered into the park. These eight visitors were probably friends with different preferences for the park facilities. Similarly, we can find groups of visitors that always visited the same locations, but never communicated. They were probably different groups but followed the same route in the park.

4 CONCLUSION

In this paper, we present a visual analytics tool, named *Spectrum*, to analyze the movement and communication data from the VAST Challenge 2015. With the help of *Spectrum*, we can identify the group of visitors who worked with the soccer star, cluster visitors with similar activity patterns into the same group and discover unusual activity patterns that are probably relevant to the incident. The intuitive interactions in *MoveView* enable users to replay and track the movements of visitors. The filtering options in *SpectrumView* help to focus on certain types of records during analysis. Numerous sorting and clustering options in this view assist in identifying groups with similar patterns. We hope that the visualization and interaction techniques demonstrated in this paper will help the analysis of other types of large scale log data.

REFERENCES

[1] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, Oct. 2008.