ParkVis: A Visual Analytic System for Anomaly Detection in DinoFun World

Abishek PURI   Dongyu LIU   Shaoyu CHEN   Siwei FU   Tianyu WANG   Yeuk-Yin CHAN*   Huamin QU†

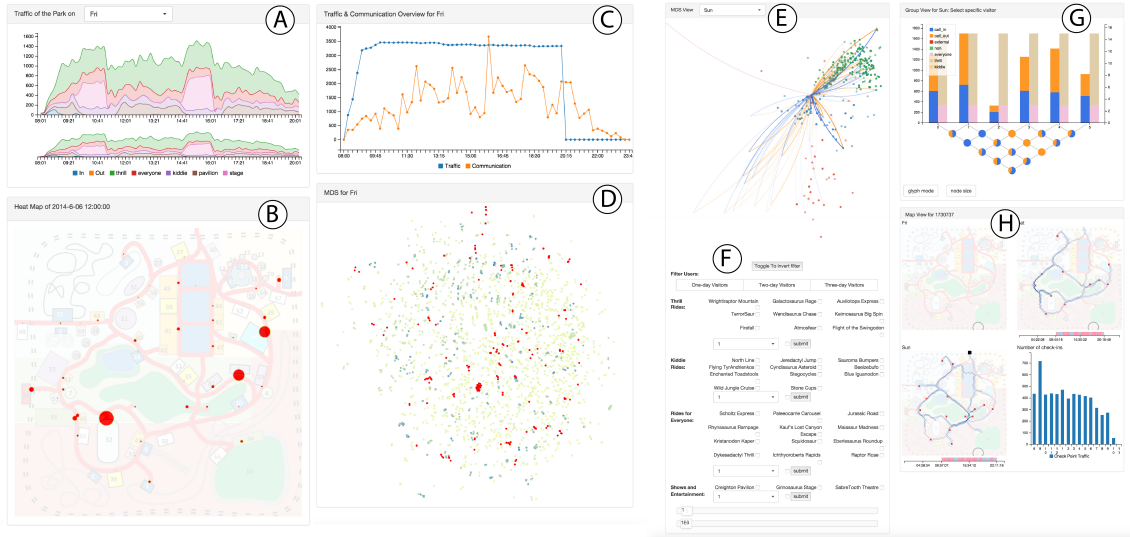The Hong Kong University of Science and Technology

Figure 1: Overview of ParkVis. A): Stacked Area Chart showing the traffic of different facilities in one day. B): Heat Map of the park in every 1 minute time stamp. C): Line Chart showing population and communication in the park. D): Multidimensional scaling (MDS) view of the whole population focusing on the movement information. E): MDS view of the whole population focusing on facilities usage. F): Filters for filtering the population in MDS based on travelling records. G): Statistics of selected groups from the MDS View. H): Path view of a selected person and the respective statistics of facilities on the day.

ABSTRACT

In this report, we propose ParkVis, a visual analytic system for tracking the unusual patterns of all paying park visitors. Using both communication and movement data, park officers can use our system to identify and gauge the extent of unusual activity occurring in the park.

1 INTRODUCTION

Exploring movement tracking information and communication data for all of the paying park visitors can bring us lots of insights. However, nowadays many tools only focus on doing data analysis to deliver numerical summaries or predictions. Understanding the group sizes, preferences of entertainment, usage of facilities and paths for the whole journey of different types of people are not easy tasks by using existing analytic tools. In this report, we propose ParkVis, an interactive visual analytic system, for park management to investigate the movement and communication data in the park. The system has various views, which helps summarize common movement patterns of visitors and find unusual patterns.

---

*e-mail: apuri, dliuae, schenan, sfuaa, twangad, yychanae @connect.ust.hk

†e-mail: huamin@cse.ust.hk

2 SYSTEM PIPELINE

Our system has two main workflows to find patterns. The first workflow mainly uses the MDS View with various filters, focusing on the check-in data of the visitors. The filters (Figure 1(F)) used are mutually exclusive by using combinations of logic controls, namely 'and', 'or', and 'not'. Analysts can use these filters to find patterns of interest. For example, analysts can inspect groups that have not played kiddy rides, potentially isolating the adult groups. Such patterns are outliers in our MDS view. We can further investigate these patterns based on their paths (Figure 1(H)) and check-in records (Figure 1(G)). Once the outliers have been identified, we can use the MDS view to see the communication data of the outliers (Figure 1(E)), potentially identifying groups that are working together. Another example is using the filter to remove all visitors who went on a thrill ride, by far the most popular class of ride. By using this method, we found a lot of unusual patterns, one of which was the entourage of Scott Jones. When inferring the group types, we focus on the cluster size and communication patterns of that cluster in the MDS view. For example, in the family group, most of the time only 1 or 2 members use the communication function of the park app. By checking the in-group communication, we can decide whether it is a family group or not. The system also shows the communication with other groups in the MDS view so that we can know whether the group has subgroups. Sometimes we will filter out groups smaller than a certain number, so that we can analyze the behavior of the large groups.

The second workflow uses spatial and temporal filters to identify unusual patterns. In this method, we first check the traffic monitor (Figure 1(A)) to choose a certain time. Then we investigate a location on the map where we can use the mouse to brush a square area in the heatmap(Figure 1(B)). All the people in the square area at that time will be highlighted in the MDS view (Figure 1(D)), allowing us to see how many people were in that location at that time. By tracking clusters of people over time, we can see whether they are in a group or coincidentally at the same location. In the MDS view, if we click on a point, the system will go to the group view (Figure 1(G)) which focuses on the check-in records so that we can know what types of visitor the point represents and know the visitor communications. This feature can help us identify and inspect suspects when we specify the venue and time.

3 SYSTEM OVERVIEW

In this section, we selectively introduce those views that showcase this design split choice that we made, as well as showcases the critical roles played by these tools in solving the grand challenge.

3.1 MDS View

The system applies the MDS algorithm [1] to cluster the population in the park. We found out that there are 42 different check-in points in DinoFun World that fit into several types, e.g. entrances, rides, entertainments (like stage, theater, Pavilion), and so on. From the Auxiliary Files, We can know that there are 3 different types of rides (Thrill Rides, Kiddie Rides and Rides for Everyone) in Dino-Fun World. We think that different types of people will have different preference for rides. Therefore, we decide to use the check-in numbers at each types of check-in points of that person to calculate the similarity of visitors using Euclidean distance. In the MDS view, a point can represent an individual or a group of people with identical check-ins, as they will share the same resultant vector from our MDS implementation. People who seldom play rides are in the top-right corner of the MDS view, and people who play a lot of rides are in the bottom-left corner of the MDS view. Also, we use different colours to indicate how many days the visitor has visited the park.

To facilitate the understanding of the characteristics of each dot in the MDS view, users are allowed to filter or highlight them by:

1. Selecting the number of facilities in each category to filter the population. For example, users can filter out visitors who did not go for a single thrill ride, or they can select visitors who went on "Galactosaurus Rage".

2. Using sliders to filter by number of overlapping points or by the total communication of a visitor. These filters are mutually exclusive by design, preventing any crossover in analysis. For example, you can filter out all users with less than 10 communication data points.

To view the communication data of a particular point, you can click on that point and curved lines will be drawn to all the senders and receivers of communication that include the selected point.

3.2 Communications Monitor

To create a visualization for our higher level view of communications in the park, we aggregated the data given to us in MC2 for each set time interval, creating a stack area chart, with the x-axis being the time in the middle of the interval and the y-axis being the number of communications broadcast in that interval. We also added a hover visual showing the aggregate value at that point. This tool was primarily used to help find communication peaks and troughs, as well as compare the day to day communication trends to see if any of the days had any anomalous communication patterns in terms of aggregate volume over given time interval.

3.3 Individual Inspector

We wanted to create a way to see individual visitors paths across the 3 days, as well as compare their position with the traffic around that position. This way, we would be able to identify if an individual happened to be in a densely or sparsely populated area in the park at that time. To do this, we created 3 map views onto which a blue line is drawn, signalling the path of the visitor taken on each of the three days. If the visitor doesn't come on a particular day, that map is left blank. We also added a bar chart, showing the population of various checkpoints around the map.This can then be compared with the position of the visitor to see if he is following the crowds or not.

3.4 Group Inspector

We wanted to create a way for the user to instantly compare all traits of visitors who have been communicating with each other. These traits include how many rides they have visited, and how much communication they had with other people. To do this, we used a stacked bar chart, with each stack representing how many times they were seen at a checkpoint. This way the user can quickly compare the traveling paths of people who are all communicating with each other, helping them understand why the communication pattern is presented this way and potentially find a new type of communication group.

We also wanted to be able to see what percentage of communications between two people was sent by each of them. To do this, we created a communication graph (Figure 2) that connects every person shown in the group Inspector. Then, we have a pie chart at the node that shows what percentage of the total communications was by person A, then doing the same for person B. This way, we can see if there is a particular person who initiates a high number of calls or a low number, giving us more information in our visual analytic system to find anomalous behavior with.
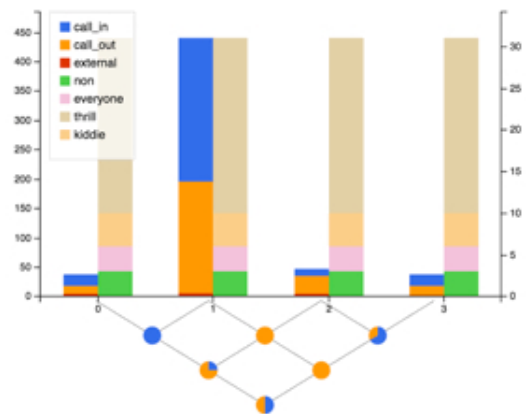

Figure 2: The Group Inspector and Communication Graph of a selected group in the MDS View

4 CONCLUSION

ParkVis provides the user with the ability to effectively filter the park visitors population and analyze both general and individual movement data. With the help of ParkVis, we can identify different types of visitors and analyze their behaviors in the park, so that the park facilities can be improved to feed the visitors need. ParkVis also allows the park management to find and analyze unusual events during operation. We hope ParkVis can play an important role in park management by integrating various kinds of data in the future.

REFERENCES

[1] T. F. Cox and M. A. Cox. *Multidimensional scaling*. CRC Press, 2000.