ITEA—interactive trajectories and events analysis: exploring sequences of spatio-temporal events in movement data
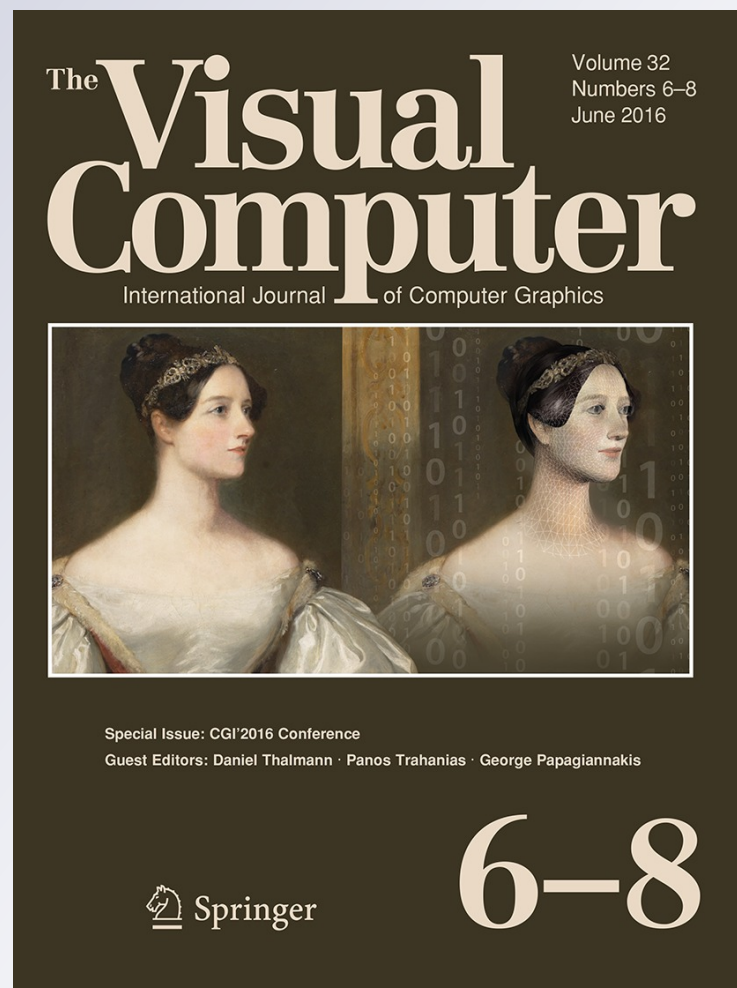
Lena Cibulski, Denis Gračanin, Alexandra Diehl, Rainer Splechtna, Mai Elshehaly, Claudio Delrieux & Krešimir Matković

Springer

Springer

CrossMark

ORIGINAL ARTICLE

ITEA—interactive trajectories and events analysis: exploring sequences of spatio-temporal events in movement data

Lena Cibulski[1,2] · **Denis Gračanin[3]** · **Alexandra Diehl[4]** · **Rainer Splechtna[1]** ·
Mai Elshehaly[5] · **Claudio Delrieux[6]** · **Krešimir Matković[1]**

Abstract Widespread use of GPS and similar technologies makes it possible to collect extensive amounts of trajectory data. These data sets are essential for reasonable decision making in various application domains. Additional information, such as events taking place along a trajectory, makes data analysis challenging, due to data size and complexity. We present an integrated solution for interactive visual analysis and exploration of events along trajectories data. Our approach supports analysis of event sequences at three different levels of abstraction, namely spatial, temporal, and events themselves. Customized views as well as standard views are combined to form a coordinated multiple views system. In addition to trajectories and events, we include on-the-fly derived data in the analysis. We evaluate our integrated solution using the IEEE VAST 2015 Challenge data set. A successful detection and characterization of malicious activity indicate the usefulness and efficiency of the presented approach.

Keywords Interactive visual analysis · Movement data · Spatio-temporal data · Coordinated multiple views

✉ Lena Cibulski
lena.cibulski@st.ovgu.de

1   VRVis Research Center, Vienna, Austria

2   University of Magdeburg, Magdeburg, Germany

3   Virginia Tech, Blacksburg, VA, USA

4   University of Buenos Aires, Buenos Aires, Argentina

5   University of Maryland, Baltimore, USA

6   University of the South, Bahía Blanca, Argentina

1 Introduction

Technological advancements make it possible to collect large amounts of movement data. For example, many commercial vehicles are equipped with position logging devices. Passenger cars will follow suit in the near future.

People also can be tracked, especially in areas such as shopping malls, railway or subway stations, or entertainment parks. Movement data help experts from various domains. For example, data can be used to prevent traffic jams, to enhance infrastructure planning, to identify critical spots, or to devise better evacuation plans. Movement data are usually large and complex and have distinct spatial characteristics. As a consequence, visual analytics is a valuable and proven tool for more effective and insightful movement-data analysis [3,5]. Movement data are increasingly augmented with additional information. When studying taxi movement data, additional information could include number of passengers, fares, or driver information. Although such data do not represent movement directly, it can significantly help in analysis. For this reason, analysis methods have to be improved accordingly.

In this paper, we deal with complex movement data that were provided for the IEEE VAST 2015 challenge [18]. The story takes place in an entertainment park equipped with an RFID tracking system. Only raw data—more than 25 millions of logs—are available. Besides movement, the visits to rides or other facilities represent important additional information. We call such kind of data events along trajectories. A large number of challenge entries and variety of proposed solutions prove the relevance of the problem and a need for novel solutions, which advance the current state-of-the-art.

We present interactive trajectories and event analysis (ITEA)—a comprehensive solution to the problem of analysis of events along trajectories. In our interactive visual

🖄 Springer

Fig. 1 An exemplary configuration of the analysis tool. **a** Initial setup of event timelines. **b** Only visitors on Sunday are brushed. **c** Refinement of brush using movements tracked. **d** Further refinement using low rides count. **e** The potential criminals remain. **f** Corresponding event timelines. **g** Graph view showing the criminals' trajectories

analysis solution, we employ coordinated multiple views (CMV) with linking and composite brushing. During the data processing step, analysts often do not know exactly which data derivatives will be needed. On-the-fly data derivation seamlessly integrates data derivation in the analysis loop, and analysis is performed without interruptions.

We classify different methods for visual analysis, where we differ between direct and aggregated methods. Additionally, we focus on analysis of event sequences, and provide several solutions based on the specific tasks. We differentiate between visualization of event sequences where time (starting time and duration of an event) is considered, and where only the order of events is important. Accordingly, we call them timed event sequences and ordered event sequences. Special attention is paid to the scalability of the proposed solution. Besides views for events along trajectories, we extensively use standard views, such as scatterplot or histogram, all linked and integrated in a single CMV tool. We evaluate the solution using the VAST challenge data.

In this paper, we are not concerned with introducing novel visualization techniques. Our focus is on leveraging the capabilities of CMV in a forensic analysis setting. This focus is motivated by the need for on-the-fly analysis in a post-attack data triage setting, in which experts seek to perform time-critical analysis tasks and obtain immediate results. Therefore, our contributions can be summarized as follows: (1) a classification of the interactive visual analysis of events along trajectories data. (2) ITEA—an integrated visual analytics framework which supports the presented

analysis approach. Different levels of aggregation are supported to cope with data size and complexity. The user can switch between different means of the analysis anytime, all in an integrated framework.

Figure 1 shows a snapshot from an analysis using ITEA. Just a small subset of available views and interactions are depicted.

2 Related work

Event sequence data appear in a very diverse range of domains, such as medical care, movement analysis, marketing, sociology, bioinformatics, or genomics, for instance [11, 12, 16]. An example of an application dealing with such event sequences is LifeFlow [19] visualization tool, where events are grouped by certain characteristics and the average of a measure (e.g., elapsed time) is calculated. The tool shows how many times a given type of event occurs and follows the sequences of events. It also provides highlighting functionality for events and subsequent events to track certain patterns in the event sequence. It does not scale well.

EventFlow [14] is an extension of LifeFlow that allows the visualization of large amounts of information. The tool filters the data, then groups different types of events and re-orders the events to show groups of sequences. Although this approach solves the cluttered and packed visualization, it hides the details of the events that compose a group. We use a similar approach allowing the user to group

and re-order sequences but we keep individual sequences. DecisionFlow [7] is a visual analytics tool designed for high-dimensional temporal event sequence data. It is scalable and provides a dynamic management of data structures, but it only handles the temporal aspects of the event sequences.

The genomic domain often deals with sequences analysis. There is usually no time or spatial aspect in genomic sequences. Sequence surveyor [1] represents an example for such an analysis. It allows for the visualization of large genomic sequences in a scalable alignment visualization. Although our work also uses sequence alignment visualization and sorting of sequences, this tool only tackled one dimension of the problem. Meanwhile, our tool presents an organization with different levels of aggregation.

There is extensive work in the area of visual analytics of movement data. An excellent survey on the state of the art in information visualization is provided by Liu et al., who give insights into current advances and challenges [10]. A recent book on analysis of movement data gives a comprehensive coverage of the topic [3]. There are several papers which deal with trajectories of vehicles. Giannotti et al. [6] propose an approach for human mobility discovery through the analysis of GPS data with M-Atlas. Guo et al. [8] present a visualization system to analyze large amounts of spatial-temporal multi-dimensional traffic trajectory data. Ferreira et al. [5] present a system that supports visual exploration of taxi traffic. They provide a visual query model that allows users to quickly select data slices and explore large amounts of spatio-temporal data. Orellana et al. [15] deal with automatic analysis of visitor movement patterns in natural recreational areas.

IEEE VAST 2015 challenge [18] submissions provide some interesting approaches. The best comprehensive solution [17] employed a web system that combines data wrangling, trajectory analysis, network analysis, and interactive visualizations for discovering movement and communication patterns of users and their networks. Buchmuller et al. [4] applied automated clustering techniques to support the analyst in exploring the data to eventually enhance situational awareness in complex analysis scenarios. Xu et al. [20] combined traditional visualization parallel coordinates, force layout, and matrix to collect and correlate data. Although many building blocks of our solution are not innovative themselves, we combined them all in an interactive framework. The user can switch between different means of analysis anytime, and the system provides views to support different tasks simultaneously.

3 Problem description

The IEEE VAST 2015 challenge [18] data include movement/location data generated by tracking the visitors of an entertainment park (DinoFun World). The park covers an area of 500 by 500 m and the visitors are tracked with a five meters resolution. In other words, the park is divided into a 100 by 100 grid where each cell has dimensions of five by five meters. The data are available for three days: Friday, Saturday, and Sunday.

The raw data are a simple list of log events. The log event contains an event description, unique visitor ID, coordinates, and a time stamp. The event description is either "check-in" or "movement". The check-in indicates that person just checked-in into a ride in the park. In addition to the log data, a park map is provided.

The rides are thematically grouped, so there are thrill rides, kiddie rides, and rides for everyone. The person is not tracked during a ride. When a visitor completes the ride, tracking continues. In addition, there are many places where visitors can shop or eat. These facilities do not explicitly log users as they enter. However, if a visitor spends a long time (we empirically set the threshold to 20 s) at the same position (five by five m cell), it is reasonable to assume that the visitor stayed at a nearby facility.

During the weekend (Saturday/Sunday), an incident happened when a local star and international soccer celebrity, Scott Jones, was visiting the park. During the visit, someone vandalized a pavilion exhibiting Jones' memorabilia, and stole an Olympic medal and possibly other irreplaceable items.

To illustrate how visual analytics can support investigation, the challenge defines three tasks: (1) characterize the park attendance at DinoFun World on this weekend. Describe up to twelve different types of groups at the park on this weekend. (2) Are there notable differences in the patterns of activity in the park across the three days? Please describe the notable differences you see. (3) What anomalies or unusual patterns do you see? Describe no more than 10 anomalies, and prioritize those unusual patterns that you think are most likely to be relevant to the crime. Each of the tasks contains additional questions [18].

As we are interested in visitors behavior, we transformed data so that each visitor gets one data record. In contrast to common visualization tools, which allow scalars only to be basic items of a record, we use a more complex data model [9]. In addition to scalar values (numerical and categorical), we allow trajectories with events and time series as basic items. The raw data contain more than 25 million log events for the three days. The final processed data contain 11,374 records which correspond to the same number of unique visitors. Figure 2 depicts the data processing steps and the structure of the complex records.

There are many possible attributes to generate. For example, one could compute a count of thrill rides per person, or number of rides at a specific attraction. As the number of options for such queries is huge, and we do not know which attributes are needed during an analysis session, we allow
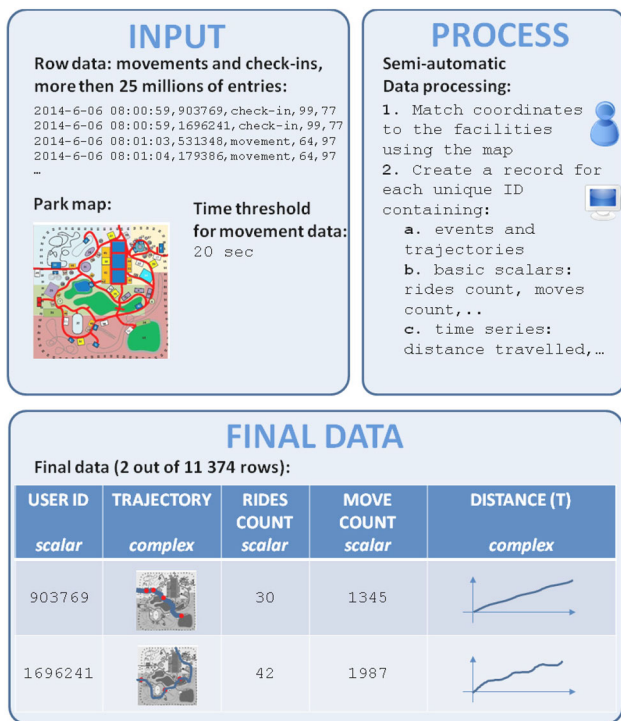
Fig. 2 Semi-automatic data transformation (*top-right*) of a simple list of log events (*top-left*) to one data record per visitor. Records consist of scalar values as well as trajectories and time series (*bottom*)

for on-the-fly data generation. This means that an analyst can request computation of a new attribute based on existing scalar and complex attributes at any time during the analysis. The session does not end, the analyst simply gets additional attributes in the pool of attributes, and can visualize them. The flexibility of the data generation significantly improves efficiency. At the same time, the analysis can start faster, as only basic attributes have to be defined in the data processing step. The pressure on analysts to define all necessary attributes in advance disappears.

Due to the characteristics of the described tasks, we have chosen to create a visitor-centric data set. The same principles can be used to create a facility-centric data set. Such a data set would be better suited to explore the park organization itself, to support optimization of park infrastructure, for example.

4 Interactive visual analysis of events along trajectories

Analyzing complex movement data is a challenging task that depends on application domain. Due to the data complexity, a pure automatic solution is often not possible. The basic tasks, as described above, are defined at a high level and it is not easy to precisely formalize detailed tasks which are needed for automatic analysis. A visual approach is of a great help here.

For example, it is much easier to interpret a trajectory when seen in the context of all trajectories rather than based on numerical data only. An analyst can easily see if a trajectory is unusual, but it is extremely hard to formally describe what makes a trajectory unusual.
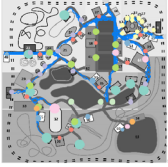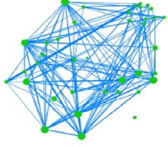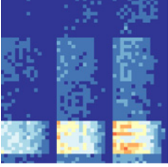
We classify visualization methods for events along trajectories using different criteria, such as consideration of time, events, and trajectories, as well as scalability. The classification of all visualization is depicted in Table 1. Each of the categories supports different tasks, and no category alone is sufficient for a complete analysis of events along trajectories data. The analyst typically simultaneously employs views from all categories. There is no predefined order how views or categories are used. Some of the views provide better overview of a large number of trajectories, while other provide more details, but do not scale well. During an analysis session, the analyst constantly switches focus between views from different categories. Therefore, for a comprehensive analysis, it is crucial to provide all views in an integrated analysis environment, and to allow a simultaneous usage of different views.

The categories of views support different analysis tasks. As we deal with spatio-temporal data, the natural starting point in the analysis is to use views that depict spatial relations. Temporal context is added in the form of animation and aggregation along time. These views give a good overview of data, and can scale well. The map, the events graph, and the heat map from Table 1 represent such views. They give an overview of the spatial relations of the trajectories and corresponding events. This allows for a basic understanding and the detection of overall trends and correlations. Usually, they are used as a starting point in the analysis. They are also used after drill down, to show the spatio-temporal characteristics of a subset of the data. During analysis, we always come back to this kind of views.

Next, in order to analyze visitors' behavior, and in order to detect patters in visitors' behavior, the analyst is interested in events that took place along a path of a movement. These tasks are supported by time lines and ordered event sequences. To better focus on events, these views do not consider spatial relations, instead, the order and duration of events are of central importance. Reducing the importance to the order only makes it easier to detect patterns of behavior. If, for example, two visitors take exactly the same rides, in exactly the same order, but with different duration of some rides, views for ordered sequences will show the same visualization for both visitors. Depicting them differently would be confusing, if we want to see if there is a pattern in the rides order. On the other hand, if we want to see if the visitors take the rides together, the timing becomes crucial.

Finally, derived data, such as scalar aggregations or derived functions of time, make it possible to efficiently drill down. For visualization of derived data, we employ stan-

Table 1 The classification of main views for events along trajectories. None of the individual views is sufficient but all views together create a powerful analysis framework

Visualization	Properties	Remark
	Map Time: ✗ Events: ✓ Traj.: ✓ Scale: 10^2	Time can be limited to an interval for deep analysis. Animation is also supported.
	Events Graph Time: ✗ Events: ✓ Traj.: ✗ Scale: 10^5	Time is handled as in map.
	Heat Map Time: ✗ Events: ✗ Traj.: ✓ Scale: 10^5	Time is handled as in map. It is possible to filter out cells with too many or too few visitors.
	Temporal Heat Map Time: ✓ Events: ✓ Traj.: ✗ Scale: 10^5	Vertical division per event gives a certain spatial aspect.
	Timelines Time: ✓ Events: ✓ Traj.: ✗ Scale: 10^3	Sequences of events and timing are in focus.
	Event Sequences Time: ✗ Events: ✓ Traj.: ✗ Scale: 10^4	Each sequence is shown in the same length—timing not in focus.
	String Sequences Time: ✗ Events: ✓ Traj.: ✗ Scale: 10^2	Easy querying—(sub)string comparison, no timing.

dard views that simplify the selection of subgroups with outstanding characteristics. Imagine that the analyst wants to see trajectories of users with a low number of rides only. She can simply derive the count of rides aggregate on-the-fly, and use a histogram, to visualize the count. Now, a simple brush in the histogram will select the desired subgroup.

All views use a consistent color scheme, when depicting events from different event categories (such as thrill rides, shows, or shopping, for example). The following sections describe each category in more detail.

5 Spatio-temporal analysis of movement data

The obvious choice when analyzing movement data is to show the trajectories themselves in a spatial context—a map view (Table 1, row 1). Such a visualization is easy to interpret. A user-controlled animation of the trajectories provides insight into the temporal formation of the movement. It works fine for few trajectories, but as the trajectories count increases, clutter and over-plotting make the visualization less and less useful. During an interactive drill-down, as the number of trajectories decreases, the map becomes more and more expressive again. The analyst can interactively filter the data in the map view based on spatial characteristics or based on time. She can also filter the data based on any derived data or sequences characteristics in corresponding views.

In addition to the trajectories, the map view also shows events. The events, the rides or visits to other facilities, are depicted using circles with sizes that correspond to the number of event occurrences. In this way, the biggest circles can easily be related to the most popular rides. The color of the circles corresponds to the event group. The animation and temporal controls influence the events visualization, too.

The map view does not scale well. As the number of trajectories increases, it is not clear what is going on. Aggregated views provide a solution to this problem since, for them visual complexity does not change as the number of visitors increases. We provide aggregated views focusing on spatial component, on temporal component, and on event sequences. Again, the user can state a time interval as well as step back and forth within the animation, for all views. Of course, these views are also fully coordinated with all other views. If an analyst brushes something in any view, the brushed data points will be highlighted in all other views.

The well-known heat map technique helps us to depict aggregated data on spatial aspects of trajectories. We color code a map cell based on the total number of visits to the cell (Table 1, row 3). To support outlier detection and detailed analysis, we allow for selective visualization based on a threshold. The analyst can specify to show only cells having the total number of visits below or above a certain threshold.

When it comes to temporal characteristics of the data, we use a temporal heat map view (Table 1, row 4). Since there is no spatial component, we arrange the map as a table. Each row corresponds to an event (or to a group of events), and each column represents a time interval. The analyst can easily see the most popular events and the temporal distribution of users across events.

6 Event sequences analysis

The event sequences represent a key feature of the given data. They are the key difference from the standard movement data, and they provide answers to questions posed in the analysis. We introduce a hierarchical approach to event sequences visualization. It supports events sequences analysis, independent of the actual path which a visitor took between two events.

At the highest level, we take the events positions into account. An events graph view, which shows the order of visited rides in the spatial context, is provided (Table 1, row 2). The events' circles are connected using straight lines whose line width corresponds to the number of visitors that have taken the two rides one after the other. Instead of the line width, a color code could be used.

Although data aggregation functions very well for spatial and temporal component, the events graph suffers from overplotting. The events keep their position (which is desired for some tasks), but due to many lines the view becomes illegible. Further, one cannot follow the whole sequence. Individual sequences cannot be reconstructed from the graph. To make the event sequences legible, we do not visualize the spatial component at the next levels in the hierarchy. The second level deals with timed event sequences (the timing of events is important here), and the third level deals with ordered event sequences (only order of events matters).

6.1 Timed event sequences

As we have a lot of sequences, we have to take a special care of scalability. One possibility to show sequences is to use a special case of a curve (Fig. 3, left). We assign *y* coordinate to each event, and *x* coordinate represents time. A density mapping is used to cope with overlapping graphs. Adjusting the line transparency and width allows for a better perception of a subset of curves. For better differentiation

and assignment of curve sections to areas, we provide a colored underlay. Colors indicate the membership of a specific event to its corresponding area. This view is highly scalable because it requires the same amount of screen space independent of the number of visitors displayed. At the same time, it is complicated to compare curves and to follow a single curve if too many curves are shown.

Figure 3, right, depicts event timelines, an approach similar to timelines used in Middguard [2], that show each sequence individually, one below each other. Movement is represented by a line, while events are displayed as colored rectangles. The colors correspond to the group to which an event belongs. Due to a large number of different events (more than 70), we decided to assign colors to the groups. For example, there is a color for kiddie rides (yellow), and all events that belong to kiddie rides (Wild Jungle Cruise, Beelzebufo,...) are shown using the same color. Information on specific event is provided by interaction, on mouse over a tooltip with the event description is displayed. Due to their representation, event timelines are easier to read than curves described above, but, at the same time, they require more screen space.

If we reduce the height of events rectangles to one pixel only, the event timelines become event colorlines. The idea is similar to the original colorlines used to depict curves [13]. The width of a line section indicates the duration of an event, while the color again corresponds to the area of an event. Colorlines use the space more efficiently than event timelines at the cost of readability.

Colorlines and event timelines can be sorted according to a specific time (Fig. 3, right). This results in sequences that represent the same events at that time being displayed below each other. This allows for investigation of visitors that took the same ride at the given time step. We also provide a slider that can be used to scroll through all samples and indicates the subset of colorlines or event timelines currently being displayed.

Fig. 3 The curve view (*left*) displays trajectories as time series using absolute times. *Colors* indicate an event's area. Event timelines (*right*) represent individual sequences. They can be sorted according to a specific time

6.2 Ordered event sequences

If we are interested only in the chronological order of events and not in their duration, we use ordered event sequences. They can be visualized using event timelines and colorlines. Individual event sections are of the same size now. The width of a section does not correspond to time anymore. We also provide merging of identical ordered sequences. We depict them only once, and show the number of visitor sequences contained in one representation. This is referred to as the count. Figure 4 shows ordered event timelines. Individual sequences for every visitor are shown on the left, while identical sequences are summarized into one representation on the right. Using these unique sequences, we can significantly decrease the number of sequences that need to be displayed, from 11,374 to 3936 for all three days, for example. Brushing based on the count of unique sequences is also supported. The analyst can select all ordered sequences having a certain number (or a range of numbers) of visitors. The visitors are brushed now, and the timed sequences' views show us if the events have same timelines as well. It is easy to identify groups of visitors in this way. When using unique sequences, the coloring works the same way as before while sorting does not use events and instead considers the count of the sequences and sorts in descending order.

If a group or a family visits a park, it is not unusual that some members of the group skip some rides. To support searching for sub sequences in the ordered event sequences, we introduce string representation. Each ordered event sequence is represented by a composition of substrings which stand for the corresponding events (Fig. 5, top-left). Each event is either represented by a detailed substring, such as T01 (which means thrill ride number one), or a group substring, such as T (which means any thrill ride). These string representations can now be used to search for a specific course of events. The analyst can query for a search term consisting of characters and digits. The asterisk character, '*', serves as a wildcard operator. The analyst can also choose which category is to be searched, for example, only rides. The query results in a brush containing all event sequences that contain the subset of ordered events as defined in the

Fig. 4 Identical sequences (*left*) are combined into uniques (*right*), reducing the number of sequences from 11,374 to 3936
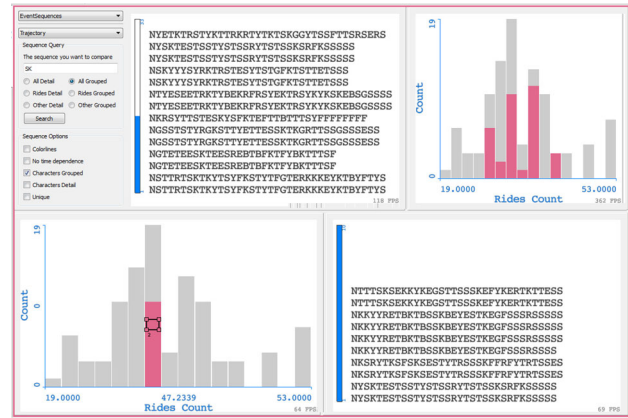

Fig. 5 String representation used to query events with a specific substring (*top-left*). Histogram updates accordingly (*top-right*). The analyst refines the query by brushing the histogram, only middle bin is selected (*bottom-left*). The string view is updated accordingly (*bottom-right*)

query. Just as for any other brush, all other views are updated accordingly as can be seen in Fig. 5.

7 Derived data analysis

In addition to the above-presented customized views focusing on interactive analysis of events along trajectories, we employ standard views, such as histogram, scatter plot, parallel coordinates, or curve view. These views are used to show aggregated values, and often serve as a basis for interactive drill down. The main task is to understand groups in the data, to find visitors with similar behavior. Timed and ordered event sequences can certainly help, but for a quick drill down aggregated values and standard view are much more efficient. If we, for example, depict kiddie rides count and thrill rides count using a scatter plot, a cluster of visitors that checked in into a lot of thrill rides but did not visit any kiddie ride is easily visible. We can brush the cluster now and explore the trajectories as well as event sequences for the selection. Of course, we can drill down further, and refine the brush in any other view.

We use numerous scalar attributes during analysis: the count of rides and of food facilities visited, the total way traveled, and so on. All of these attributes can be computed on-the-fly, during the analysis. These attributes enable us to study the impact of patterns concerning event characteristics on numerical attributes. Besides scalar attributes, we also derive the distance traveled as a function of time. For each visitor, this graph shows the cumulative way she has traversed during the park visit. Figure 1b, e shows such a graph. Visitors' movement patterns such as speed and walking/resting are clearly visible. Furthermore, visitors can be easily compared by means of visiting hours, agility, resting time, speed, etc. It is also possible to identify and brush

groups with similar movement patterns for further investigation. We can detect outliers such as late comers and follow the graph line to investigate their movement. Outstanding and, therefore, potentially suspicious behavior, for example a combination of long rest periods and unusual visiting hours, can be efficiently revealed using the distance as a function of time. Brushing of such visitors highlights the corresponding graph line and allows for an interactive detailed analysis of additional attributes using linked views. This view is also highly scalable as its required screen space is not dependent on the number of visitors displayed.

8 Evaluation

We evaluate the newly proposed approach to analysis of events along trajectories on the VAST Challenge data set. Due to space limitations, we provide just a few examples, which show how we solved the challenge tasks. We start the analysis with an overview using the temporal heat map (Fig. 6). It clearly shows the peak times for certain events. We see that entries are mostly active in the first hour, between 8:00 and 9:00, and that the top entrance is the most popular one. We also see that thrill rides are the most often visited rides in the park. The visits to the pavilion and to the stage have clear peak times, which indicate the times when Scott Jones was there. After the first overview is gained, we proceed with the analysis.
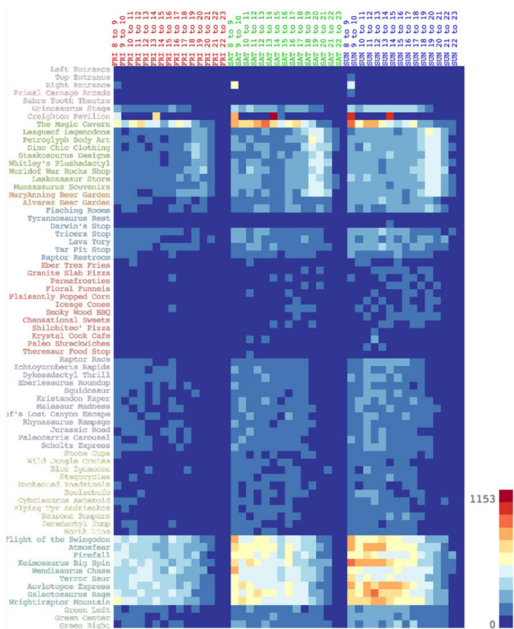
Fig. 6 The temporal heat map view shows the temporal aggregation of data. The *y* axis shows park facilities, and the *x* axis shows time in 1-h interval

One of the tasks asked to identify different types of visitor groups. Unique sequences are well suited for this question, because they directly represent groups of visitors that moved through the park together. Most of the visitors are organized in small groups of 2–11 persons. The remaining visitors almost evenly split up into large groups and single visitors. Brushing according to different visitor counts leads to large groups consisting of 29–44 visitors, smaller groups with 2–11 members and 1569 out of 11,374 visitors that discover the attractions on their own.

Analyzing the walked distances as function of time for these groups with the help of the curve view reveals interesting patterns. Single visitors predominantly enter the park shortly after the opening (8:00–10:00). Very few enter at mid-day (13:00–15:00), but none of them late afternoon. They are very eager: besides entering early, nearly half of the single visitors also visit the park on all three days. Most of the visitors in smaller groups enter the park early in the morning. Some of them enter at mid-day and others late afternoon (17:00–18:00). To both of the previously named groups applies: each day there are a few visitors that leave the park at mid-day and do not come back. On Sunday, this number is highest, which might be caused by people having to go to work on Monday and traveling back home Sunday afternoon. There is no large group that stays for more than one day. Interestingly, all the 30 large groups enter the park right after the opening. They even leave the park late in the evening, making the most out of the available time. This might be due to time scheduling of big tourist groups.

Another task was to describe unusual patterns that may be related to the crime. By brushing the curve view for late-coming visitors, we detect an outlier. He is the only visitor who had come late on Friday and was there on Saturday as well. Additionally, he moved very little and only visited four attractions including the entry. Examination of the visitor's event timeline shows that he did not leave the park on Friday. Instead he took one very long ride in the night from Friday to Saturday. Taking a look at the trajectory, we can see that he first got something to eat at Chensational Sweets (39), then directly went to the Alvarez Beer Garden (33) and stayed there for over four h. It is likely that the visitor got more or less drunk in the meantime. After leaving the Beer Garden, he checked in to Scholtz Express (20) and did not move until the next morning. At 08:00:27 on Saturday, the individual moved again and directly took the same entry he came in to leave the park. From this movement pattern, we can reason that the individual was probably drunk and found a place to sleep in the ride area of Scholtz Express. He then was awakened by the park's opening on Saturday morning and hurried to leave the park. Querying for the detailed string representation of this individual using ordered event sequences reveals that he in fact was not alone but member of a five person's group. Switching to event timelines illustrates that only one of them,

the drunk visitor, took a longer last ride than the others because he fell asleep. Finally, analyzing the corresponding trajectories clearly shows that they together checked in at 21:53:10, but only four of them left after half an hour.

Searching for rides that took a conspicuously long time may lead us towards potential criminals that left their tracking device somewhere to sneak around the park. Figure 1 shows an overview of the interactive drill down. Using timed event sequences, we detect six individuals who stayed in a restroom for the second half of a day. They can be recognized by the wide blue bars in the event timelines displayed in Fig. 1a. As they only visited the park on Sunday and did not move for multiple hours, we brush Sunday visitors in curve view (Fig. 1b) as well as low movements tracked and low rides count in histogram (Fig. 1c, d). The resulting number of brushed visitors can be well distinguished in curve view with walked distances so we can identify and brush the lines corresponding to the suspects as can be seen in Fig. 1e. All views are updated according to the combined brushes, resulting in only the six potential thieves being displayed using timed event sequences in Fig. 1f. Examining the trajectories of the suspects allows for a deeper insight of the movement. Figure 1g depicts the spatial relations of the events that the suspects visited as graph view. Simultaneously using the curve view for timed events helps for the temporal placement of the current trajectory animation step. The individuals entered the park at 08:48:45 and spent about 20 min there, possibly making last agreements on their plan. Afterwards, they spent 7:45 min at the Creighton Pavilion (32) where Scott Jones' memorabilia were exhibited, maybe for preparation of the future theft. In the following time, they visited an unremarkable combination of kiddie and thrill rides as well as rides for everyone. They also made a detour to shopping attractions twice. At 14:55:25, they checked in to Tyrannosaurus Rest (66) which is located right beside the stage where Scott Jones performs. They did not move for almost six h until 20:41:30. It is very likely that they left their tracking devices in the restroom, where they could take them off without observation. Using the absence of security guards, who were needed at the stage during a part of the time period, they could commit the crime without being tracked or interrupted. They then left the park without returning to get their devices, which is why the tracking stops in the restroom. Only for one of the six suspects, the tracking system registers another check-in at a shopping store after the restroom stay. Possibly he was chosen to deposit the captured items there, as it would be too dangerous to get them out of the park right away.

When investigating the curve view with walked distance as function of time, one could ask if similar movement patterns of visitors relate to similar timelines or other characteristics being alike. Brushing individuals with similar visiting times and distances walked allows for further investigation of additional attributes. Scatter plot view reveals visitors with different characteristics. Most of them visit at least a few thrill rides. But there are visitors that at the same time do not visit a kiddie ride at all, while others do attend multiple kiddie rides. When focusing on kiddie rides and shopping, it becomes clear that there is one subgroup that likes to go shopping but does not visit any kiddie rides while other visitors attend both categories of attractions. Using these linked views for analysis leads to the conclusion that similar movement patterns not necessarily mean similar characteristics of additional attributes.

9 Conclusions and future work

The analysis of spatio-temporal events in movement data is a challenging task. Due to data and task complexity, a pure automatic analysis solution is not sufficient. We conquer data complexity using interaction, data aggregation, and a hierarchical approach to event sequence analysis. We provide a classification of views which all focus on different aspects of the data and, therefore, allow for visual analysis from different perspectives. Each of the views alone would not be sufficient to analyze the data, but all of them integrated in a unified framework offer great possibilities. There are so many different ways to compute additional data, that it is practically impossible to come with a complete list of needed data prior to analysis. Therefore, we support on-the-fly data derivation, which was extensively used. We were not aware of data complexity when we started to analyze the VAST Challenge data. The solution presented in this paper is a result from numerous discussions and trials and errors. We have started with the spatio-temporal views, and as we ended in a dead end, it became clear that this was simply not sufficient. We plan to further pursue this research and to additionally focus on the ride-centric approach. Reorganizing data is straight forward, and all concepts introduced can easily be applied. We see a great challenge in a simultaneous analysis of both approaches. Integrating them would open new, unmatched, analysis possibilities. To make it possible, novel analysis procedures, and novel interaction and visualization design will be necessary.

References

1. Albers, D., Dewey, C., Gleicher, M.: Sequence surveyor: leveraging overview for scalable genomic alignment visualization. IEEE TVCG **17**(12), 2392–2401 (2011)
2. Andrews, C., Billingsy, J.: Middguard at DinoFun World. VAST Challenge 2015 Entry (2015)

3. Andrienko, G., Andrienko, N., Bak, P., Keim, D., Wrobel, S.: Visual analytics of movement. Springer (2013)

4. Buchmuller, J., Fischer, F., Streeb, D., Keim, D.A.: Using visual analytics to provide situation awareness for movement and communication data. In: Proceedings of the 2015 IEEE VAST, pp. 121–122 (2015)

5. Ferreira, N., Poco, J., Vo, H.T., Freire, J., Silva, C.T.: Visual exploration of big spatio-temporal urban data: a study of New York City taxi trips. IEEE TVCG **19**(12), 2149–2158 (2013)

6. Giannotti, F., Nanni, M., Pedreschi, D., Pinelli, F., Renso, C., Rinzivillo, S., Trasarti, R.: Unveiling the complexity of human mobility by querying and mining massive trajectory data. VLDB J **20**(5), (2011)

7. Gotz, D., Stavropoulos, H.: DecisionFlow: visual analytics for high-dimensional temporal event sequence data. IEEE TVCG **20**(12), 1783–1792 (2014)

8. Guo, H., Wang, Z., Yu, B., Zhao, H., Yuan, X.: TripVista: triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In: 2011 IEEE PacificVis, pp. 163–170 (2011)

9. Konyha, Z., Matković, K., Gračanin, D., Jelović, M., Hauser, H.: Interactive visual analysis of families of function graphs. IEEE TVCG **12**(6), 1373–1385 (2006)

10. Liu, S., Cui, W., Wu, Y., Liu, M.: A survey on information visualization: recent advances and challenges. Vis. Comput. **30**(12), 1373–1393 (2014)

11. Malik, S., Du, F., Monroe, M., Onukwugha, E., Plaisant, C., Shneiderman, B.: Cohort comparison of event sequences with balanced integration of visual analytics and statistics. In: Proceedings of the 20th international conference on intelligent user interfaces, IUI '15, pp. 38–49. ACM, New York (2015)

12. Mannila, H., Toivonen, H., Inkeri Verkamo, A.: Discovery of frequent episodes in event sequences. Data Min. Knowl. Discov. **1**(3), 259–289 (1997)

13. Matkovic, K., Gracanin, D., Konyha, Z., Hauser, H.: ColorLines view: an approach to visualization of families of function graphs. In: 11th International Conference on Information Visualisation, IV 2007, pp. 59–64 (2007)

14. Monroe, M., Lan, R., Lee, H., Plaisant, C., Shneiderman, B.: Temporal event sequence simplification. IEEE TVCG **19**(12), 2227–2236 (2013)

15. Orellana, D., Bregt, A.K., Ligtenberg, A., Wachowicz, M.: Exploring visitor movement patterns in natural recreational areas. Tour. Manag. **33**(3), 672–682 (2012)

16. Patnaik, D., Butler, P., Ramakrishnan, N., Parida, L., Keller, B.J., Hanauer, D.A.: Experiences with mining temporal event sequences from electronic medical records: initial successes and some challenges. In: Proceedings of the 17th ACM SIGKDD. KDD '11, pp. 360–368. ACM, New York (2011)

17. Steptoe, M., Krueger, R., Zhang, Y., Liang, X., Garcia, R., Kadambi, S., Luo, W., Ertl, T., Maciejewski, R.: VAST challenge 2015: grand challenge–team VADER/VIS award for outstanding comprehensive submission. In: Proceedings of the 2015 IEEE VAST, pp. 119–120 (2015)

18. Whiting, M., Cook, K., Grinstein, G., Fallon, J., Liggett, K., Staheli, D., Crouser, J.: VAST challenge 2015: Mayhem at Dinofun World. In: Proceedings of the 2015 IEEE VAST, pp. 113–118 (2015)

19. Wongsuphasawat, K.: Guerra Gómez, J.A., Plaisant, C., Wang, T.D., Taieb-Maimon, M., Shneiderman, B.: LifeFlow: visualizing an overview of event sequences. In: Proceedings of the SIGCHI conference on human factors in computing systems. CHI '11, pp. 1747–1756. ACM, New York (2011)

20. Xu, J., Ren, S., Tao, Y., Lin, H.: A collaborative visual analysis system for communication pattern discovery. In: Proceedings of the 2015 IEEE VAST, pp. 139–140 (2015)

Lena Cibulski received her bachelor's degree in computational visualistics from the Otto-von-Guericke University Magdeburg in 2016. She is currently studying for her master's degree in computer science. Her research interests are in the fields of information visualization and scientific visualization, with special focus on interaction techniques as well as visual analysis of heterogeneous data that include complex data types and cohort study data.

Denis Gračanin received the BS and MS degrees in electrical engineering from the University of Zagreb, Croatia, in 1985 and 1988, respectively, and the MS and PhD degrees in computer science from the University of Louisiana at Lafayette in 1992 and 1994, respectively. He is an associate Professor in the Department of Computer Science at Virginia Tech. His research interests include virtual reality and distributed simulation. He is a senior member of ACM and IEEE and a member of AAAI, APS, ASEE and SIAM.

Alexandra Diehl is a PhD student at the University of Buenos Aires, Argentina, since 2008. She received her diploma degree in Computer Software Engineering from the University of Buenos Aires (UBA), Argentina, in 2005. Since 2009, she is a teaching assistant in the course of Algorithms and Data Structures I at the Faculty of Exact and Natural Sciences, UBA. Since 2016, she is a teaching assistant in the course Information Visualization at the Data Mining Master of the UBA. Her research interests include visual analytics, visualization techniques for geospatial data, and environmental sciences.

Rainer Splechtna graduated in computer science at Vienna University of Technology in 2003. He is currently working as a researcher at VRVis Research Center in Vienna. His research interests include virtual and augmented reality, scientific visualization, information visualization, and visual analytics with special focus on interaction techniques and visualization and analysis of multivariate time-dependent data.

Claudio Delrieux is full professor and PI at the Electrical and Computer Engineering Dept., Universidad Nacional del Sur, Bahia Blanca, Argentina. He holds a BSEE (1990), MSCS (1996), and PhD in CS (2004), all from the Unversidad Nacional del Sur. He was Fulbright Posdoctoral Fellow at the University of Denver during 2006 and 2007. His research interests include Image Processing, Scientific Visualization, and Computational Intelligence.

Mai Elshehaly received a PhD degree and a Master degree in computer science from Virginia Tech, where she worked at the Center for Human Computer Interaction, advised by Prof. Denis Gracanin. She is currently working as a postdoctoral research associate at University of Maryland, Baltimore County (UMBC), where she joined the Interactive Visual Computing Lab (IVCL) and is advised by Prof. Jian Chen. Her research interest is in scientific data visualization, with special focus on interactive rendering and large graph comparison for time-variant data analysis.

Krešimir Matković is a senior researcher and the leader of the Interactive Visualization Group at VRVis Research Center in Vienna. He is interested in extending visual analysis technology to challenging heterogeneous data, in particular to a combination of multi-variate data and more complex data types, such as functions, for example. He also focuses his research on developing a structured model for visual analysis which supports a synergetic combination of user interaction and computational analysis. He teaches at TU Vienna and at the University of Zagreb. He is a member of ACM, Eurographics, and IEEE Computer Society.