STAD-HD: Spatial Temporal Anomaly Detection for Heterogeneous Data through Visual Analytics

Yu Zhang[1]* Guozheng Li[1] Chufan Lai[1] Qiangqiang Liu[1] Shuai Chen[1] Lu Feng[1]
Tangzhi Ye[1] Siming Chen[1] Ren Zuo[1] Zhuo Zhang[2]† Zhanyi Wang[2] Xin Huang[2]
Fengchao Xu[2] Li Yu[2] Shunlong Zhang[2] Qiusheng Li[2] Xiaoru Yuan[1]‡

1) Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University
2) Qihoo 360 Technology Co. Ltd.

Abstract

The capability of providing situation awareness in the scenarios with complex data environment is critical and challenging. Especially, it is hard to quickly identify patterns and anomalies from multiple heterogeneous data sources with spatiotemporal and high-dimensional information. To solve the problem, we propose STAD-HD, an event-based visual analytic system focused on semi-automatic anomaly-detection that links heterogeneous time-varying data with spatial and temporal filtering. We evaluate STAD-HD with mini challenge 2 and 3 of VAST Challenge 2016, and identify valuable patterns and anomalies with the system.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical user interface

1 Introduction

Despite the abundance of visualization systems and mining algorithms for time series and trajectories, it is hard to efficiently explore a spatiotemporal dataset with numerous time-varying objects and ambiguous task specifications. Take the VAST Challenge 2016 dataset as an example. In this dataset, there are around five hundred time series collected from building sensors distributed at different positions, and around a hundred trajectories of ID cards generated from fixed and mobile sensors with different spatiotemporal resolution. Besides, part of the data is received in a streaming manner. Dealing with the broad task of pattern, anomaly, and relationship detection, analysts can be easily burdened by the mass of details.

STAD-HD is a web-based system aimed at this problem. It exploits the idea that the burden to browse original information space can be reduced by starting from checking highly possible anomalies. According to the classification metric in [1], this can be regarded as an event-based approach. With the perspective of anomaly detection, it is important to enable users to explore suspicious events interactively, relate heterogeneous time-varying data with spatiotemporal filtering [2], and drill down to details.

To support these desirable functionalities, we propose a pipeline (shown in Figure 1) for the implementation of STAD-HD. In Data Preprocessing step, trajectories are generated from ID card data. In Anomaly-based Filtering step, anomalies in the preprocessed data are semi-automatically detected. Users can apply the spatiotemporal information of anomalies for cross filtering to drive the change of views in the Sensor Readings Component and Trajectory Component in Visualization step. The spatiotemporal range of anomalies can be iteratively specified in the visualization for filtering.

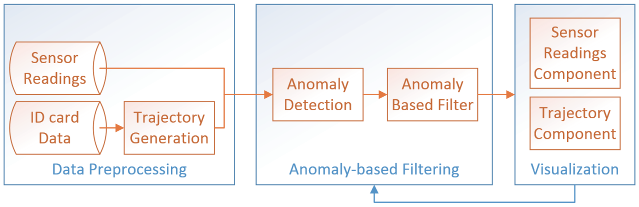


Figure 1: Pipeline of STAD-HD.

2 Anomaly Detection

Two mechanisms are used to detect anomalies in building sensors and trajectories respectively listed as follows.

For the building sensor data, we apply a heuristic to detect anomalies: a timestamp is abnormal when the reading goes beyond a threshold defined by the average and standard deviation of readings at a given time point in all days. Notice that statistical results are calculated separately for weekdays and weekends, since the patterns of weekdays and weekends have apparent differences.

For the trajectories, we define two kinds of anomalies. One kind of anomalies is that an employee spends a long time in a zone without his own office or public areas (e.g. meeting room). The other kind is that the record of fixed and mobile ID card sensors conflicts, namely someone appears at different places at one timestamp.

In the system, all the anomalies are explicitly displayed in a Warning Stack View in text form, and sorted in chronological order. We define the data format for all kinds of warning entries uniformly as <temporal tag, spatial tag, attribute>.

3 Visual Analytic System

STAD-HD contains two major components: Sensor Readings Component and Trajectory Component.

The two components are linked with a shared map and timeline for spatial and temporal cross filtering. Users can also select an event entry in the Warning Stack to conduct spatiotemporal filtering with the event's spatial tag and temporal tag.

To support real-time surveillance and replay of streaming data, we enable users to follow or rewind the stream at any ratio, which is equivalent to setting timestamp of interest periodically.

The major components are introduced in the following.

3.1 Sensor Readings Component

Sensor Readings Component (shown in Figure 2) leverages three views: Position-centric Glyph View paired with Attribute-centric Glyph View for situation awareness at one timestamp, and Line Chart List View for browsing individual attributes.

In Position-centric Glyph View, a node denotes a zone, a link reveals physical connectedness, and a fan out of glyph on the node represents a building sensor attribute with radius and color encoding normalized sensor reading. By comparison, in Attribute-centric

*e-mail: yuzhang94@pku.edu.cn
†e-mail: zhangzhuo@360.cn
‡e-mail: xiaoru.yuan@pku.edu.cn

Glyph View, a node denotes a building sensor attribute, and each fan out of glyph on the node represents a position.

These two views show the readings of all the attributes at all the positions at one timestamp. The dual visual design of position-centric and attribute-centric glyphs facilitates fast recognition of both the events that an attribute is abnormal at many positions, and the events that many attributes at one position are abnormal.

Line Chart List View consists of line charts derived from the product of attribute selection and position selection. In every line chart, some possibly abnormal substrings are highlighted in red.

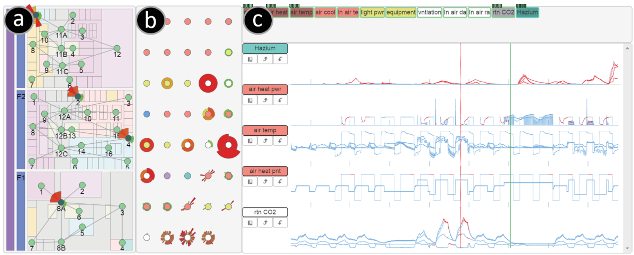


Figure 2: Sensor Readings Component: (a) Position-centric Glyph View; (b) Attribute-centric Glyph View; (c) Line Chart List View.

3.2 Trajectory Component

To visualize the overview+detail of employees' trajectories, Trajectory Component (shown in Figure 3) mainly exploits two views: Trajectory Animation View and Gantt Chart List View.

In the Trajectory Animation View, trajectories of all the employees on a selected floor are displayed with animation. A circle on the map represents an employee with its color denoting department. Anomalies are highlighted with yellow and red flash. When the exact position cannot be ascertained by the provided data, we use a strategy for location inference: if the employee is in the zone containing his office, we assume that he is in the office; else if the employee is in a zone with public areas (e.g. meeting room), we assume that he is in a public area; else his position is randomly assigned in the corridor. Different inferences are represented with different border styles of the circle.

As for the Gantt Chart List View, the discrete visiting sequences of all the employees are visualized in a static form, covering all the timestamp as an overview. We also provide a detailed Gantt Chart for a single person (shown in Figure 4b), in which a short period of stay is represented with a circle, making it easily observable.

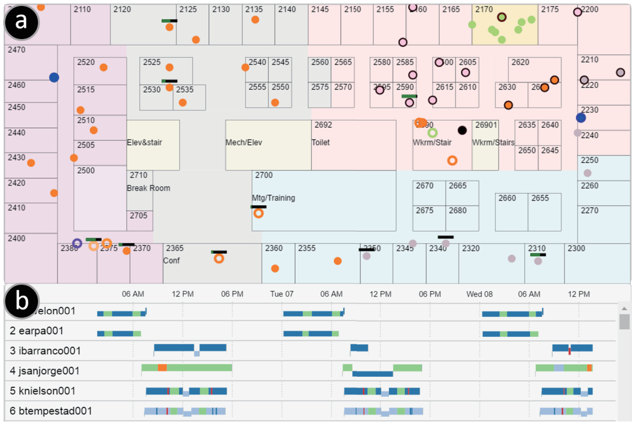


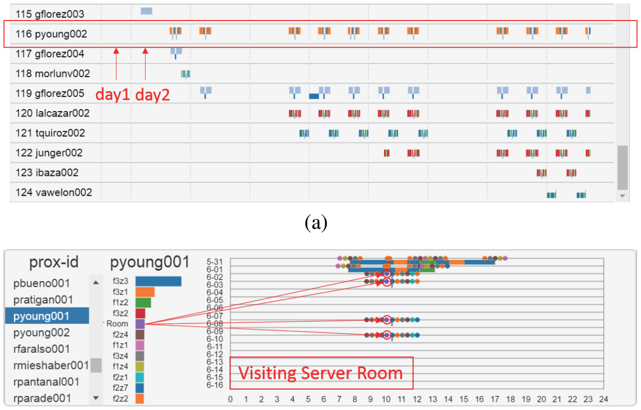
Figure 3: Trajectory Component: (a) Trajectory Animation View; (b) Gantt Chart List View.

4 Case Study

In the following, we describe an anomaly-driven exploration case with STAD-HD. We start from browsing the Gantt Chart List View, and find the ID card "pyoung002" not used in the first two days of the dataset (shown in Figure 4a), which implies that its owner lost his old ID card and got this card as a duplicate. In this list overview, it is easy to identify that less than ten people lost their ID cards in the time range concerned in the dataset, and therefore this rare event deserves further inspection.

Then, we go for the lost ID card named "pyoung001", and find that oddly enough, this card was still in use after the original owner got a duplicate. More importantly, after the owner got a duplicate, the old card was especially used for accessing the Server Room, presumably a room for setting the HVAC system of the building (shown in Figure 4b). We form the assumption that the card was used by someone other than the former owner.

With the Gantt chart of pyoung001's trajectory, we can reveal the fact that it was detected around **Room 2345**, the office of **Bennett Loretta**, which suggests that Loretta stole this card and used the card to access the Server Room.



(a)

(b)

Figure 4: (a) pyoung002's trajectory in Gantt Chart List View; (b) Detailed Gantt Chart of pyoung001's trajectory.

5 Conclusion

To meet the challenge of handling numerous time series and trajectories, we focus on semi-automatic anomaly-detection and developed a system named STAD-HD with an underlying pipeline. We use VAST Challenge 2016 dataset for case study to demonstrate the effectiveness of the system, which enables users to start exploration easily from abnormal events.

Acknowledgements

The authors wish to thank IEEE VAST Challenge committee and reviewers. This work is supported by NSFC No. 61170204, and partially funded by NSFC Key Project No. 61232012 and the National Program on Key Basic Research Project (973 Program) No. 2015CB352500. This work is also supported by PKU-Qihoo Joint Data Visual Analytics Research Center.

References

[1] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 14(1):47–60, 2008.

[2] S. Chen, X. Yuan, Z. Wang, C. Guo, J. Liang, Z. Wang, X. Zhang, and J. Zhang. Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):270–279, 2016.