A Collaborative Visual Analytics of Trajectory and Transaction Data for Digital Forensics

VAST 2014 Mini-Challenge 2：Award for Outstanding Visualization and Analysis

Ying Zhao[1], Yanni Peng[1], Wei Huang[1], Yong Li[1], Fangfang Zhou[1], Zhifang Liao[1], Kang Zhang[2,3]

1 Central South University; 2 Tianjin University; 3 University of Texas at Dallas.

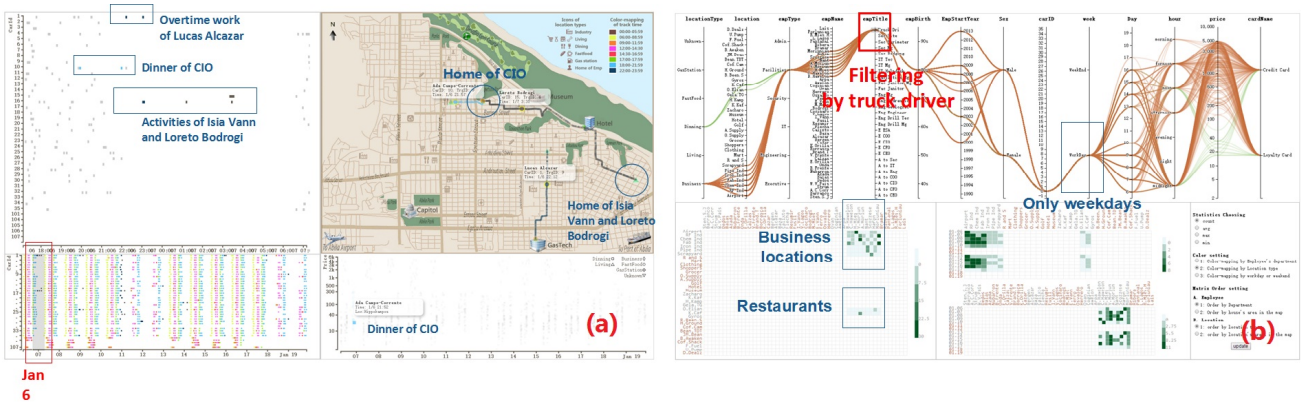

Figure 1: (a) The SGGViz tool: the anomalies at midnight of Jan 6; (b) The PMViz tool: the transactions of truck drivers.

ABSTRACT

Advanced digital forensics technologies provide powerful basis in criminal investigation. Due to the complexity and diversity of data as well as the increasing quantity of the data, traditional digital forensics technologies have already can not adapt to the analysis requirements. This paper provides a visualization approach to analyze multiple types of data for digital forensics which provide users three interrelated tools: the RadViz tool, the PMViz tool and the SGGViz tool. Our solution focuses on the correlation analysis of trajectories and transactions, and it plays an important role in the process of analyzing the case in VAST 2014 Mini-challenge 2.

Keywords: Digital forensics, visual analysis, trajectory data, transaction data, spatial and temporal analysis, parallel coordinates, matrix, RadViz

1 INTRODUCTION

With the widespread popularization and application of the computer and network, crime cases involving computers increase year by year. Criminals may leave many traces in the form of electronic data during the criminal activities [1]. Investigators use types of digital forensics tools to extract the electronic data from computers and networks, and check out clues and evidences to reveal the truth of crimes. As a newly emerged and multi-disciplinary research field, visual analytics offers friendly interactions and connects the communications between human and data. Therefore visual analytics can establish a strong mental model based on the capability of human cognition to find out the hidden information from data during digital forensics, especially when the type of data is various [2].

This paper provides a visualization approach to analyze multiple types of data from VAST 2014 Mini-challenge 2. The Mini-challenge 2 mainly provides trajectory data, transaction data and some resumes of employees in GASTech. Our solution is particularly designed for analyzing features of trajectory data and transaction data. To address the challenge, our solution presents three tools which can help find suspicious information of track or transaction and provide data support for the analysis of the crime. These tools include RadViz clustering locations and employees in transaction data, PMViz for analyzing transactions and their statistical features, and SGGViz discerning spatio-temporal characteristics of trajectories and transactions.

2 DATA PREPROCESSING

With possible issues such as missing data, conflicting data, data of varying resolutions, outliers, or other kinds of confusing data, the original data is imperfect. Due to the uncertainties and conflicts inherent in the original datasets, we mainly conduct three aspects of processing work.

First, we cluster location types and employees by using RadViz[3] which clusters locations to identify location types, addresses the uncertainties of location types, finds employees with common consumption habits and creates a better order in matrix view. Figure 2 shows the identification of the location type of “Hippokampos” whose location type can not be identified from name and the raw data. Through the RadViz tool, “Hippokampos” clustered together with “Katerina Caf” and its records only occur at 13:00, 19:00, 20:00, 21:00, so we think that “Hippokampos” is likely a restaurant.

Second, to help users learn more about the specific stops of each car and the corresponding consumption information, we segment car-tracking data of each car by the continuity analysis of time on raw GPS data. Then the start and end of each section of trajectories can be marked by triangle and bigger circle, and the temporal features of each trajectory can be coded by different color (upper-right of Figure 1(a)).

Third, there are not the precise geographic positions of consumer places and employee's home in the official tourist map. So, we utilize the consumption records and corresponding segmented car-tracking data to locate their relatively precise positions. Finally, we create a composite map by calibrating and incorporating multiple map layers, including: tourist-map, road-network, location and legend layer to complement the map information in official tourist map (upper-right of Figure 1(a)).

{ zhaoying511, zhoufangfang}@gmail.com;
zfliao@csu.edu.cn;
kzhang@utdallas.edu

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

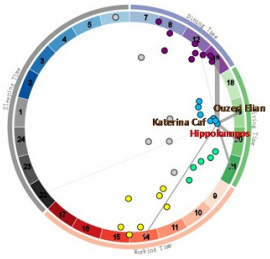

Figure 2: The RadViz tool: identify location type of "Hippokampos".

3 Visualizations

3.1 The PMViz Tool

To analyze transaction data in-depth, we designed the PMViz tool (Figure 1(b)) which consists of a parallel coordinates view and three matrix views for analyzing transactions and their statistical features. To increase the flexibility of analysis and find the hidden patterns, we adopt other related properties to extend the dimensions and use a parallel coordinates view to make multi-dimensional analysis where attributes are represented by parallel vertical axes. Each data item is represented by a polygonal line that intersects each axis at respective attribute data value. The three matrix views analyze the statistics data mainly from three dimensions: time, place and consumer (Figure 1(b)), and the statistical items include the number, average, minimum and maximum of transactions. Dimension rearrangement is also applied in this view to help users explore data.

3.2 The SGGViz Tool

The trajectory data generally contains information such as latitude and longitude, time and speed. To observe the spatio-temporal patterns of trajectories and transactions, we provided the SGGViz tool (Figure 1(a)) which consists of two scatter views, a Gantt chart and a GIS view.

With the SGGViz tool, analysts can analyze the spatio-temporal characteristics with three views. A trajectory scatter view (bottom-left of Figure 1(a)) can help analysts observe the temporal characteristics of trajectories in an overall level. By selecting data in trajectory scatter view, the selected data will be mapped to the Gantt chart automatically (upper-left of Figure 1(a)). In the Gantt chart, a trajectory is indicated by a elongated rectangular box which can represent the start/end time and duration of trajectories simultaneously. Analysts can query the corresponding staff and time by selecting a specific trajectory or car in the Gantt chart. To show the spatial distribution of trajectories, query stops of trajectories, the GIS view (upper-right of Figure 1(a)) was adopted in SGGViz. A triangle in the map represents the start point of a trajectory and a circle represents the end point.

To have a conjoint analysis with transactions, we specially designed a transaction scatter view (bottom-right of Figure 1(a)). Analysts can observe the transaction data from three aspects: price, time and the transaction place type. The X-axis represents the time, Y-axis represents the price and the shape of the points indicates the type of transaction place.

The views above supplement each other, by combining these views, analysts can observe the spatio-temporal patterns of trajectories and analyze trajectories along with transactions which can help better analyze the behaviour patterns of transactions and the movements of people.

4 Case Studies

In the following, we briefly discuss some anomalous findings in Mini-challenge 2.

In Figure 1(a), the points highlighted are the trajectories of Isia Vann (car 16) , Loreto Bodrogi (car15) and CIO on Jan 6. The GIS view shows that Isia Vann (car 16) went from his house to the place near CIO's home at about 23:00 on Jan 6 first, and then Loreto Bodrogi (car15) arrived at the same position at 03:00 on Jan 7 and they had stayed near CIO's house for the whole night. We speculate that they planned to monitor CIO by turns. Another important clue is that Isia Vann and Loreto Bodrogi are good at surveillance equipment and alarm device which is depicted in their resumes. To sum up, we highly suspect that these two men were involved in some mysterious monitoring events on executives of GASTech.

The truck drivers are special according to their activities and consumptions. They have their unique characteristics different from the general staffs. From figure 1(b) we can see the transactions of truck drivers which shows that truck drivers not only often run between business places and company but also rarely consume in restaurants which is entirely and totally different from others.

By using the mode of manual query with SGGViz, analysts can select the car and the time period they are interested in. In this case, we consider the transactions and trajectories simultaneously, we can analyze the spatio-temporal characteristics of transactions and have conjoint analysis with trajectories. In Figure 3 we can see five suspicious places and an abnormal transaction record. The five suspicious places are often visited by four suspects working in the security department without any corresponding transaction records. And a transaction of 10,000 dollar at "Autosupply" without relevant trajectories of Lucas Alcazar (car 1) on Jan 13 is strange.

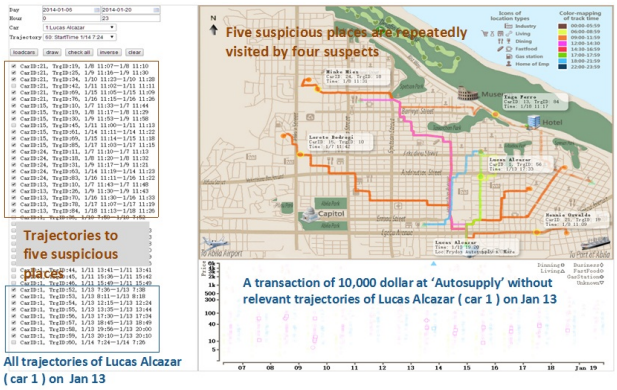

Figure 3: The SGGViz tool: five suspicious places and an abnormal transaction.

5 Conclusion

The tools in the analysis of Mini-challenge 2 played a huge role. By using these tools, analysts can observe the spatio-temporal patterns of trajectories, detect anomalies, identify suspects, locate abnormal activity locations and consumption records, and explore ways of crime and the criminal process.

6 Acknowledgements

This work is supported by the National Natural Science Foundations of China under Grant No. 61103108 and 61400511.

References

[1] Zareen M.S. Waqar A. Aslam B. Digital forensics: Latest challenges and response[J]. Information Assurance. 2013, 11: 21-29.

[2] Malik A, Maciejewski R, Collins T F, et al. Visual analytics law enforcement toolkit[C]//Technologies for Homeland Security (HST), 2010 IEEE International Conference on. IEEE, 2010: 222-228.

[3] Daniels K, Grinstein G, Russell A, et al. Properties of normalized radial visualizations[J]. Information Visualization, 2012: 1473871612439357.