**Zhi-fang Liao****·Yong Li·Yanni Peng·Ying Zhao\*·Fang-fang Zhou·Zhi-ning Liao·Sandra Dudley·Mohammad Ghavami**

**A Semantic Enhanced Trajectory Visual Analytics for Digital Forensic**

**Abstract** With the increasing application of GPS devices, trajectory data has been frequently adopted in digital forensics because it can encompass spatial and temporal aspects of the suspects’ movements. However the lack of semantic information causes the difficulty of linking the trajectories with the activities of the suspects. In the situation of a kidnapping, this paper proposes a semantic enhanced method in trajectory analysis, which categorizes the daily activities of suspects into different semantic status in time and space by connecting trajectory data with transaction data. In the meantime, we present an interactive visualization system with four inner-linked views to provide a collaborative visual analytics of trajectory and transaction data in multiple perspectives. In the case study, the kidnapping investigation is used to demonstrate how the system works on the routine pattern analysis of suspects, the detection of abnormal behaviors, and the association exploration among suspects and their abnormal behaviors.

**Keywords** Visual analytics, digital forensics, criminal investigation, trajectory data, spatio-temporal data;

# 1 Introduction

With the widespread popularization and application of computers, criminal investigation will eventually turn to computer technologybecause criminals may leave many traces in the form of electronic data during the criminal activities (Carrier et al. [2003](#ref1)). Digital forensics is the science of dealing with the evidences from computers, networks, and other electronic devices by identifying, collecting, preserving, documenting, examining, analyzing, and presenting (Casey [2011](#ref2); Lang et al. [2014](#ref3)). Over the past decade, researches on digital forensics focused on the basic problems of forensic procedures (Beebe and Clark [2005](#ref4)), the principles of the investigation, and the design and process models of digital forensic tools (Reith et al. [2002](#ref6)). With the rapid increment of the sources and volume of the data, the analysis of digital data in criminal investigation is becoming more complex. Many researchers have focused on figuring out new ways to make digital forensics more intuitive, [substantial](http://cn.bing.com/dict/search?q=substantial&FORM=BDVSP6) and vivid.

With the advanced data collection technologies, such as GPS devices and mobile phones, various types of trajectory data are increasingly complementing. Trajectory data has recently received substantial attention in the visualization community. A number of visual tools for trajectory analysis have been applied in many domains, such as helping traffic planners understand traffic jams (Guo et al. [2011](#ref7)) and assisting sociologists in analyzing the behavior of individuals (Liao et al. [2010](#ref8)).

In the meantime, trajectory data has been applied in digital forensics frequently because it can encompass the spatial and temporal aspects of the suspects’ movements (Lee et al. [2008](#ref9); Aravecchia et al. [2010](#ref10)). However the important clues are always overwhelmed in a mass of routine movement data since the sample rate of trajectory data is high. Moreover, the lack of semantic information in trajectory data brings the difficulty in connecting trajectories with human behaviors.

Ying Zhao (Corresponding Author)**·**Fangfang Zhou

School of Information Science and Engineering, Central South University, 410083 Changsha, Hunan, China

E-mail: [zhaoying511@gmail.com](mailto:zhaoying511@gmail.com)

Zhi-fang Liao • Yong Li • Yanni Peng

School of Software, Central South University, 410075 Changsha, Hunan, China

Zhining Liao·Sandra Dudley·Mohammad Ghavami

School of Engineering, London South Bank University, London, UK

This work focuses on the semantic enhanced trajectory analysis for rapid detection of abnormal events and efficient search of suspects for law enforcement. It can make the interview and field survey more pertinent, and shorten the period of investigation. On the background of a kidnapping, an interactive visualization system is designed to support a collaborative visual analytics of trajectory and transaction data for the criminal investigation. To improve the description of the trajectory, we first partition the car tracking sequence of a person into trajectory segments by extracting the start/stop information. Then, we categorize the motivations of driving into different semantic types in the semantic background of movement destinations or related transactions. Thus, each trajectory segment can be assigned with the relevant semantic information to indicate the motivation of driving. Inspired by the Gantt chart in project management, a novel Gantt view is designed to visualize the serialized activities of multiple persons and days. Meanwhile, three other views, including the timeline view, the scatter view and the GIS view, are designed in the system. Four inner-linked views provide semantic enhanced trajectory analysis from the perspectives of time, space, individuals and activities. Finally, in the case study, the kidnapping investigation is used to demonstrate how the system works on the routine pattern analysis of suspects, the detection of abnormal events, and the association exploration among suspects and their abnormal activities.

The main contributions of the paper are: (1) an interactive visualization system for digital forensics, which provides an integrated analysis of trajectory data and transaction data in multiple perspectives. (2) A semantic enhanced trajectory analysis, which provides the intuitive exploration of the human daily activities for depicting the general life routines of individuals and detecting unusual behaviors of suspects.

# 2 Related work

Digital forensics is an evolving andcomplicated field focusing on analyzing evidence from electronic devices. It begins with the establishment of the Federal Bureau of Investigation (FBI) computer analysis response group CART in 1984 (Noblett et al. [2000](#ref11); Pollitt [2010](#ref12)). In 1990s, the first international digital forensics software Expert Witness for Mac (predecessor of EnCase (Pasquale et al. [2013](#ref13); EnCase Forensics Homepage [EB/OL]. [2014](#ref14))) was born. After that, many computer forensics tools such as X-ways Forensics (X-Ways Software Technology AG[EB/OL]. [2014](#ref17); Huebner et al. [2007](#ref15);) appeared one after another. However, the existing tools have some limitations because of format incompatibilities, encryption, or even simply a lack of training (Garfinkel and Simson [2010](#ref34)).

As a newly emerged inter-discipline research, visual analytics presents the data in the form of images, offers friendly interactions, and connects the communications between human and data to find out the hidden information. Many researchers have introduced visual analytics into the domain of digital forensics and criminal investigations to help investigators extract more evidence from a mass of data. Crime map (Chainey and Ratcliffe 2005; Mburu and Helbich 2014) is one of the typical applications. For example, (Heim [2014](#ref19)) developed the visualization methodology to create the hotspot map of crimes; (Malik et al. [2010](#ref20)) proposed the VALET system to identify regions with higher probabilities of criminal incidents in West Lafayette, Indiana. Another typical application domain is the cyber security visualization (Shiravi et al. [2012](#ref21)). For example, PCAV (Choi et al. [2009](#ref22)) visualized the patterns of internet attacks, such as Internet worms, DDoS attacks and network scan, on parallel coordinates; Zhao (Zhao et al. [2014](#ref23)) presented a multi-perspective and deductive tool, MVSec, to detect traces of cyber-crimes in heterogeneous cyber security datasets.

Nowadays, one of the important topics in visual analytics is trajectory data analysis. Many visualization techniques have been implemented on trajectory data, such as the spatio-temporal aggregation (Andrienko et al. [2008](#ref24)), density maps (Scheepens et al. [2011](#ref25)), interactive Kohonen Maps (Schreck et al. [2009](#ref35)) and stacked 3D trajectory bands (Tominski et al. [2012](#ref26)). Meanwhile, the trajectory data have been used in various application domains. The example users are urban managers who need to understand the traffic jam and optimize traffic networks (Wang et al. [2013](#ref27)), security managers who need to find the better escape routes when the nuclear leakage occurs (Song et al. [2013](#ref28); Song et al. [2014](#ref29)), and sociologists who need to analyze the refugees’ migration behavior over time (Boyandin et al. [2010](#ref36)). Most of these works address only macroscopic challenges, and concentrate on the trajectory data itself (e.g. form, speed, direction). Some methods have been proposed to enrich trajectory data. For example, Andrienko (Andrienko et al. [2013](#ref31)) proposed abstract trajectory visualization based on transformation from the geographic space to an abstract semantic space to analyze human activities; Krüger emphasized overall patterns of large-scale electric vehicle trajectories during a certain time period or within a specific area in (Krüger et al. [2013](#ref30)). Further heenriched GPS-enabled vehicle trajectories with destination types (charging station, home location, etc.) through semantic POI (point of interest) information, and he also presented a visualization system for the analysis of movement behaviors (Krüger et al. [2014](#ref32)). While his visual expressions and research perspectives are not applicable to the comparative analysis on multi-person patterns simultaneously. (Kapler et al. [2004](#ref37)) proposed a GeoTime information model to visualize given events in time and space with 3-D view and analyze associations among events and relationships among people.

As a result, it can be summarized that visualization techniques are widely used not only in digital forensics but also in the analysis of trajectory data. The sense making (Pirolli et al. [2005](#ref38)) of trajectory can provide more dynamic information about criminal cases than that of intelligence. Meanwhile the semantic enrichment is one of the key issues on the trajectory analysis to explore human behaviors and detect abnormal events in criminal investigation.

# 3 The scenarios and data preprocessing

# 3.1 The scenarios

The application background of this system is a kidnapping which is provided by VAST 2014 Mini-challenge 2 (MC2 2014) (VAST Challenge Homepage [EB/OL]. [2014](#ref33)). In the roughly twenty years, GAStech has been operating a natural gas production site in the island country of Kronos. GAStech provided many of their employees with company cars for their personal and professional use. But unbeknownst to the employees, GAStech has equipped the cars with GPS tracking devices to monitor them. In January 2014, several employees of GAStech were abducted in the midst of a company celebration. Soon the kidnappers approached to ransom demands. Immediately, the law enforcement from Kronos launched the criminal investigation which started from the employees in the company. The car-tracking data and the transaction data of the employees’ personal credit cards were provided to the law enforcement to support their investigation. Data were only available for the two weeks prior to the kidnapping.

Early in the investigation of this case, the investigators were not quite familiar with the urgent situation in GAStech, and had very few clues of the case. So they wanted to gain deeper insights into the car-tracking data and the transaction data at hand, and dig up some valuable information. It could make the interview and field survey more pertinent. It would also save manpower and material resources, and shorten the period of investigation. Firstly, the investigators wanted to quickly understand the general life routines of all employees, such as: what was the time to go to work, what was the time to go off work and so on. They also wanted to master the activity routines of each employee, such as: the common areas where an employee went, the common places which an employee liked to transact at and so on. Secondly, they wanted to know what unusual events could be seen in the data and why these events were significant. The rapid detection of abnormal behaviors and the quick access to the relevant information (characters, time, trajectories, transaction places and transaction amount, etc.) would be very helpful. Finally, they needed to reason the abnormal behaviors by some questions, such as: what was the purpose of the abnormal behavior? Which abnormal events were linked internally? Whether these abnormal events were relative to the key events of the kidnapping (e.g. criminal premeditation, tool preparation and checking out the important locations)? Who were likely the victims and the suspects?

The visual analysis in this case investigation should meet the requirements as follows based on the scenarios above: (1) Allow integrated analysis of trajectory data with transaction data. (2) Quickly depict the general daily routine of all employees and the personal routine of each employee. (3) Highlight the abnormal behaviors and easily extract the detailed information of the anomalies. (4) Allow investigators to reason the inner link of several anomalies in the character, location, and time. In conclusion, the aim of this system is to assist the law enforcement to lock suspects and reason the criminal process. It makes the field investigation and the interview more pertinent, while also improves the efficiency of the criminal investigation.

# 3.2 Data preparation

The MC2 2014 has provided the car-tracking data, the transaction data, the road network and a tourist map. The car-tracking data includes the information of time, carID, person and geospatial coordinate set; the transaction data contains the information of time, person, location name and price; the tourist map only has a few places marked by icons, and there are no geographic positions of all transaction locations, employees’ home and the company. Like most datasets, the above data is imperfect, and there are two important issues need to be addressed. The first one is the uncertainty for “What locations are involved”. Thus, it is necessary to figure out the precise locations of the car parked and the transaction took place on the map. The second one is the uncertainty for “What events happened”, so we should enrich trajectory data with semantic information to support the exploration of employees’ behavior patterns. Thus, the preprocessing mainly consists of four parts: trajectory segmenttation, location classification, geographic targeting and semantic assignment.

**Trajectory Segmentation:** The raw car-tracking data is a sequence of time-ordered spatial points, where each point has a timestamp, a carID and a geospatial coordinate set. As shown in Fig.1 (a), the track points of car 23 are projected to the original map, and they are continuous path without information about where the car stopped or started. It is difficult for investigators to connect the spatio-temporal nodes with employees’ activities. Thus, we partition the sequence of car-tracking data to many trajectory segments by a threshold value which represents the time continuity. A segmented trajectory is termed a “SegTraj” in this paper. For example, when the time [interval](http://cn.bing.com/dict/search?q=interval&FORM=BDVSP6) between two consecutive points is greater than the threshold, these two data points will be respectively set as the end of the first SegTraj and the beginning of the second SegTraj. As shown Fig.1 (b), the continuous path of car 23 is segmented into trajectories with the starting points (orange dots), and ending points (green triangles).

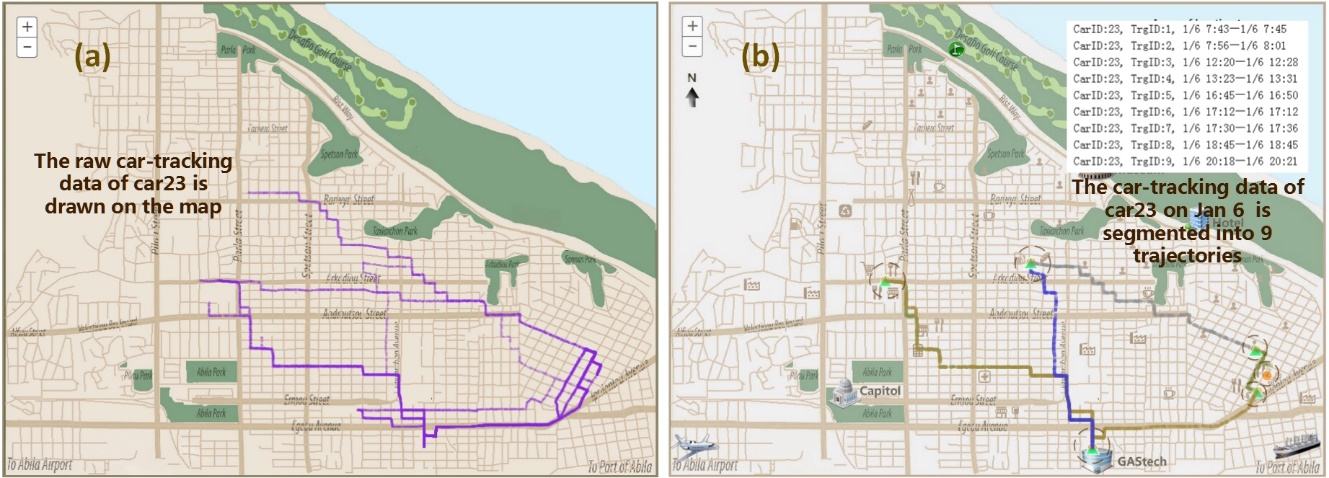


Fig.1 The segmentation of car-tracking data. (a) The raw track points of car 23 are drawn directly on the map. (b) The segmented tracking data with starting and ending points of car 23 on Jan 6.

**Location Classification:** The classification of the locations helps the law enforcement to know the motivation of driving. For example, if the destination of a trajectory is a restaurant, the motivation of the movement is dining in general. In this case, the relevant locations are classified.

1. Home location：A home location is defined as a place which a person departs from and returns to every day.
2. Working location: The working location (GAStech) is defined as the destination that is most frequently visited by all car users.
3. Transaction places: Other places of interest are extracted from the transaction data. According to the common sense and social attributes of locations, these places are categorized into six categories: Fast Food, Living, Dining, Gas Station, Industry and Unknown.

**Geographic** [**Targeting**](http://cn.bing.com/dict/search?q=targeting&FORM=BDVSP6)**:** In this step, the transaction places/employees’ homes/GAStech will be calibrated on the original map. In general, the GPS points of the same place will be close on the map. So, after the synchronization between the transaction data and the car-tracking data, each place can be targeted based on the density of GPS points. In this paper, DBScan (Ester et al. 1996), a density-based clustering algorithm is applied.

1. Home location: For a specific person’s home, the destination points of the last SegTraj of this person every day are extracted. The density center of them is defined as his/her home location.
2. GAStech company：The densest area of the destination points is the place of company. Thus, the density center of them indicates the location of the company.
3. Transaction places: The GPS point of each transaction record can be determined by the consumer and the time in the car-tracking data. The GPS points of a transaction place can be extracted according to all transaction records at this place. The density center of them is the precise position of the transaction place.

**Semantic Assignment：**In the perspective of trajectory analysis, a person’s daily life can be generally classified into two states: driving or not driving. In the crime investigation, it’s important to know the motivation of driving or what he/she has done in destination, but the trajectory data cannot directly provide the semantic information. Thus in this processing, two steps are implemented to associate the trajectory data with semantic information.

1. Semantic classification: According to the location classification, the motivation of driving or the activities of human not-driving can be categorized into eight semantic types: working, staying at home, dining, living (shopping, fixing cars in a shop, playing in a park, and so on), visiting (visiting friends or relatives), staying in a hotel, refueling and unknown.
2. Semantic enhancement: For a specific SegTraj, we add an additional attribute to indicate the motivation of driving. The value of the attribute is one of eight semantic types according to the SegTraj’s destination location type. For example, when a SegTraj’s destination is a home location, if the owner of the home and the driver of the SegTraj are the same person, the activity after this SegTraj is defined as ‘staying at home’; otherwise it could be defined as ‘visiting’.

With the data preprocessing, the track sequence of each car on the map is divided into trajectory segments with the starting and ending point. The transaction places and other important locations are located and marked out by different icons on the map. Semantic information is associated with the trajectory segments, to indicate the human activities. To sum up, the data preprocessing lays a solid foundation for achieving a smoother visual analysis on multiple datasets in the perspectives of time, space, person and semantics.

# 4 Visualization

Depending on system requirements and data preparation, a visualization system with four coordinated visual views and abundant interactions is provided to help investigators get deeper insights into trajectory data in criminal investigation. Fig.2 shows the overview of the system which includes the timeline view, the scatter view, the Gantt view and the GIS view.

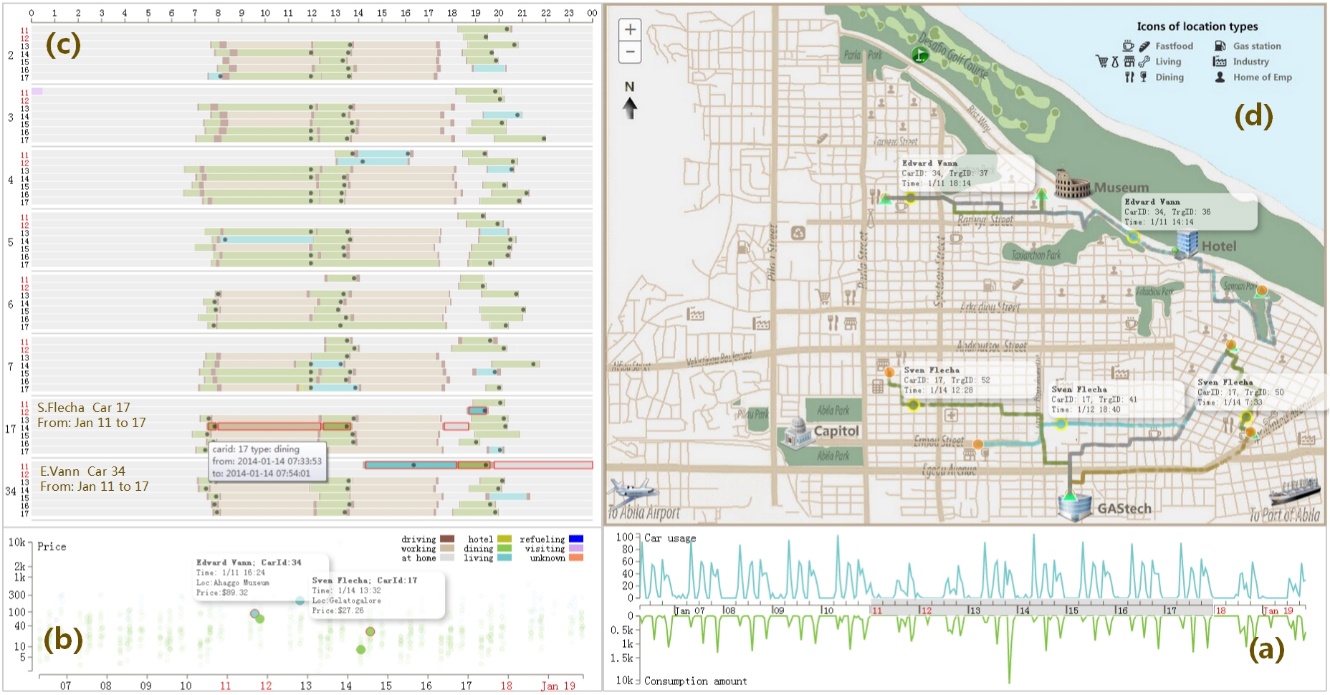


Fig.2 The system overview. (a) The timeline view presents the macro trends of statistical information extracted from trajectory data and transaction data over time from Jan 6 2014 to Jan 19 2014. (b) The scatter view with detailed information of all transaction records. (c) The Gantt view presents semantic activity sequences of eight persons, and the small dots in the bar indicate the relevant transaction records. (d) The GIS view with trajectories of car 34 on Jan 11 and car 17 from Jan 11 to Jan 17 2014.

# 4.1 The timeline view and the scatter view

The temporal information is always one of the core elements in criminal investigation, because the case development is dynamic and the case [deduction](http://cn.bing.com/dict/search?q=deduction&FORM=BDVSP6) is time-dependent. Therefore, the analysis of temporal characteristics is offered both in the timeline view and in the scatter view.

The timeline view presents the macro trends of statistical information extracted from the trajectory data and the transaction data over time. So in a way, it is not only the start of the temporal analysis but also the start of the whole work flow. To help investigators find out more implicit information from the comparison between the trajectory data and the transaction data, we adopt a “back to back” timeline diagram, in which X-axis is shared by two timelines and two Y-axis are respectively drawn upwards and downwards. In Fig.3 (a), the upward timeline shows a time series related to trajectory data (e.g. the number of track points, the number of driving cars and the average speed of all driving cars), and the downward timeline shows a time series related to credit card transaction data (e.g. the transaction times, the transaction amount). In the interactions, the timeline view works as the main filter of time in the whole system. The investigators can select their concerned time period in this view to perform a synchronized update in the other views.

The scatter view is particular designed for the transaction data. The transaction data not only is used to enrich semantic information of trajectories, but also offers a chance to achieve a more powerful investigation on transaction patterns. The view can display the distribution of price in time and the detailed transaction information. Shown as Fig.3 (b), the X-axis and Y-axis respectively represent time and price. Each colored dot in the scatter diagram indicates a transaction record, and the color of dot is related to the type of the transaction place. A click on a colored dot can unfold the detailed information of the transaction.

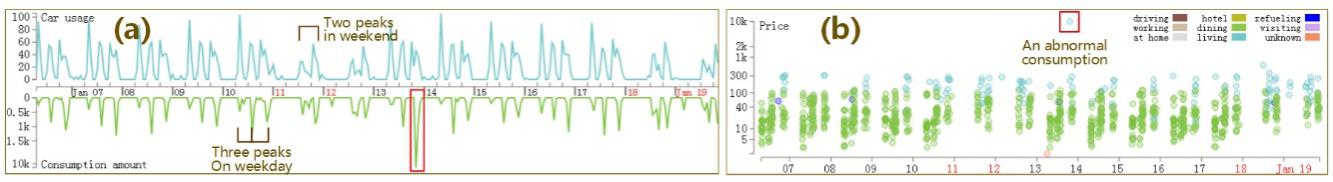


Fig.3 The overview of the timeline view and the scatter view. (a) The timeline view shows a three-step pattern on weekday and a two-step pattern in weekend. (b) The scatter view shows an abnormal transaction record which is much higher than others.

# 4.2 The Gantt view

In the data preparation, the human activities have been divided into eight semantic types, such as working, dining or staying at home. Obviously, the activity of a person in a specific spatio-temporal node is unique, so the various activities of a person in a continuous period of time can be serialized. The horizontal elongated rectangular boxes are lined up from nose to tail to represent activity sequences, shown as Fig.4 (a). As it is very similar to the Gantt chart in project management, this view in the system is called as “The Gantt View”.

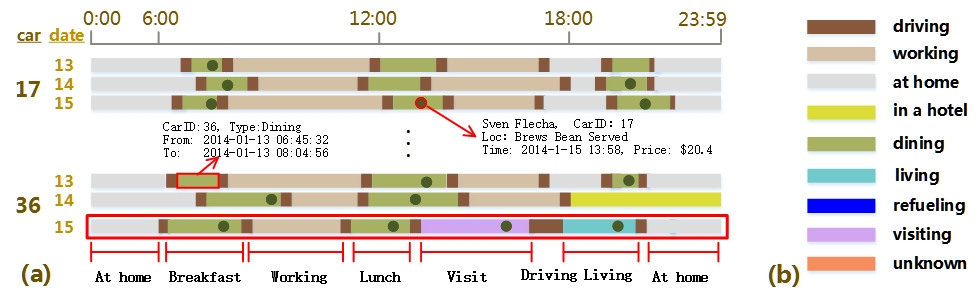


Fig.4 Schematic Gantt view and color coding. (a) The Gantt view with the activity sequences of car 17 and car 36 from Jan 13 to Jan 15 2014. (b) The color scheme of the semantic types of activities.

The Gantt view in the system can provide the comprehensive analysis of behavior patterns from multiple perspectives, such as time, persons, and semantic types. Shown in Fig.4 (a), the X-axis represents time, and the Y-axis indicates PersonId or CarID. Each row in the Gantt view represents the activity sequence of a person in a day; each horizontal bar represents an activity of a person; the length of the bar indicates the duration of the activity, and the color indicates the semantic type of the activity. The color scheme is shown in Fig.4 (b). For example silver gray is going/staying at home and brown indicates working. Furthermore, this color scheme is kept [consistent](http://cn.bing.com/dict/search?q=consistent&FORM=BDVSP6) in the other three views of the system. The small dot in a bar means a transaction record happened in the corresponding time. When multiple concerned days are selected, activities of the same person in these days will be arranged in date order.

The interactions in the Gantt view are abundant. Investigators can get details of an activity by clicking a bar. Flexible filter options about time, person and activity are also provided to investigators. In addition, when investigators select any activity in the Gantt view, the relevant trajectories and transaction information will be displayed or highlighted on the GIS view and the scatter view.

# 4.3 The GIS view

The spatial information is another core element in criminal investigation, because it plays the crucial role in assisting investigators to locate abnormal activities. The GIS view is particular designed for observing individual trajectories in days or multi-person trajectories during a certain time period. For the accuracy of criminal investigation, the GIS view in this system adopts segmented and point-to-point method to draw trajectories instead of utilizing the traditional abstract way like the density map.

Shown in Fig.2 (d), the visual trajectory in the GIS view is made up of dense car-tracking points. The triangle and circle respectively represent the starting point and the ending point of a trajectory segment. In the data preparation, each trajectory has been assigned with semantic information. Therefore, the color of semantic enhanced trajectory is consistent with the color of the activity in destination. In the interactions, the detail information of each trajectory can be obtained by clicking on any track points, as well as the corresponding transaction record of each trajectory will be highlighted in the scatter view. For example, Fig.2 (d) shows the path of Edvard Vann (car 34) on Jan 11. He first went to “Ahaggo Museum” from home at 14:10, then he went to “Hippokampos” for dinner, and finally he went back home at 19:37.

# 5 Case studies

In this section, three cases from the kidnapping investigation are carried out to illustrate how our visualization system works. These cases include the pattern analysis of all employees’ daily life, the abnormal behavior detection, and the association exploration of activities involving multiple people.

# 5.1 The routine analysis of employees’ daily life

To stand out the suspicious persons and activities, the first thing investigators should do is to master the normal behavior patterns of employees. In a macro point of view, the timeline view reveals many common [characteristics](http://cn.bing.com/dict/search?q=Characteristics&FORM=BDVSP6) of employees’ daily routines. As shown in Fig.3 (a), there are three distinct peaks of car usage respectively happened in the morning, noon and nightfall on weekdays in the upward timeline. This may mean that they drive to eat/work/go home. Furthermore, the peaks in the morning are steeper which reflects office workers’ morning is in a hurry, and the peaks at nightfall last longer which indicates that the night life is more colorful. In weekend, the features in the upward timeline are slightly different. The peaks in the morning are completely disappeared, and the peaks in the afternoon last longer. It illustrates that employees like to start their activities from noon in weekends. In the downward timeline, the peaks of transactions are basically consistent with the car usage peaks. Moreover, the lower transaction amount in the morning indicates the cheaper fast food is preferred in a hurry morning.

The scatter view in Fig.3 (b) shows that the transaction data presents a three-step pattern on weekday and a two-step pattern on weekend. These are similar to the periodical characteristics in the timeline view. Further observation finds that the entertainment and shopping transactions (blue points) significantly increase in weekends, which is also fit with the common sense.

The Gantt view combined with the GIS view can give a more intuitive explanation of employees’ daily life. Take for example Sven Flecha’s (car 17) activities in a week, Fig.2 (c) and Fig.2 (d) sketch out his normal life. On weekdays, he had breakfast near his home, went to the company before 08:30, had his noon breaks from 12:00 to 14:30, and went home after 17:30. At the weekend, he had enough time to stay at home, go out shopping, visiting or having dinner in a restaurant.

# 5.2 The anomaly detection from individual activities

Abnormities are more valuable to criminal investigations, because they can provide more clues related to crimes. In this case, a clue is found from an abnormal transaction record, and a suspect is i[dentified](http://cn.bing.com/dict/search?q=Identity&FORM=BDVSP6) with further investigation by our system.

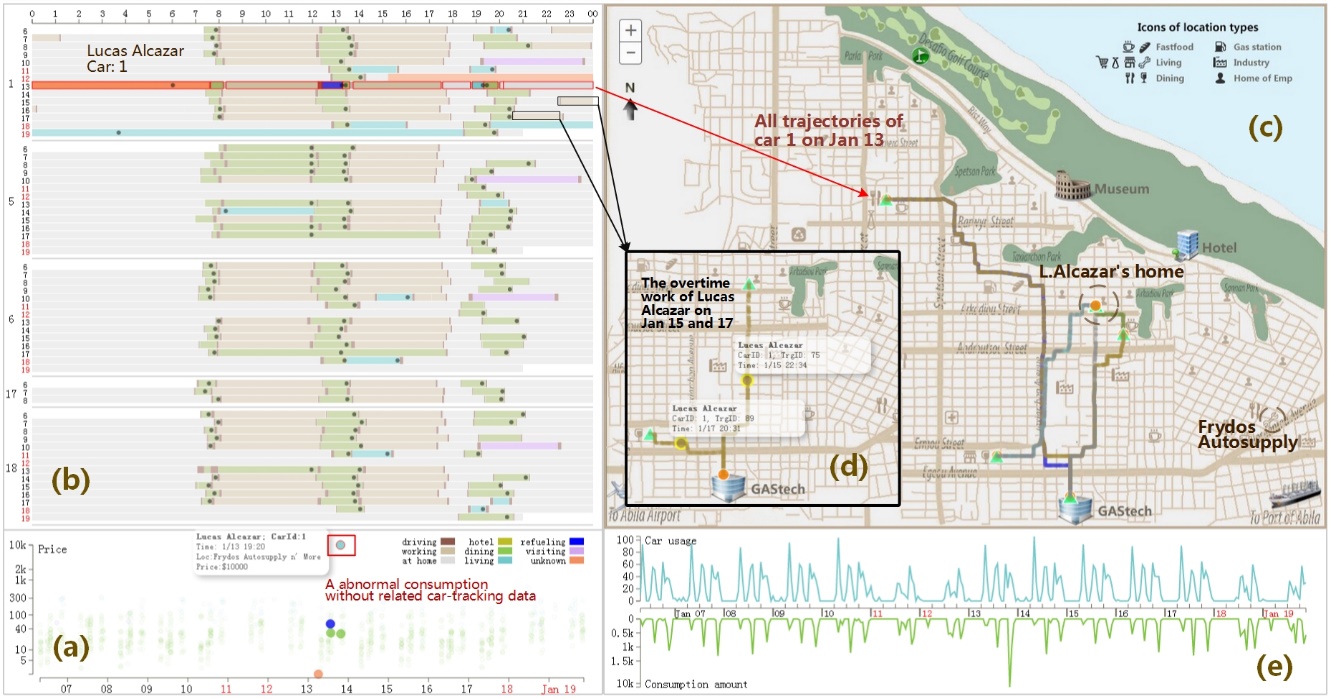


Fig.5 Lucas Alcazar's (car 1) abnormal transactions and overtime work. (a) The abnormal transaction without related car-tracking data. (b) The two-week activity sequences of Lucas Alcazar and his colleagues working in the same department. (c) All trajectories of Lucas Alcazar on Jan 13. (d) The overtime work trajectories of Lucas Alcazar on Jan 15 and Jan 17. (e) The timeline view.

The scatter view and the timeline view both show a similar obvious anomaly, shown in Fig.5 (a) and Fig.5 (e), which is a charge record of 10,000 dollar happened in “Frydos Autosupply” (an auto supply store) on Jan 13, but the average charge price of this store is only about 300 dollars. Thus, the owner of this record, Lucas Alcazar (car 1, IT Helpdesk) is listed as a suspect, and then a more detailed investigation on him is carried out.

Fig.5 (b) lists the two-week activities of L.Alcazar. And we can assume that he often works overtime in the middle of the night (as illustrated in Fig.5 (d)). But, the late-night activity rarely happened to any of his colleagues working in the same department, such as Isak Baza (car 5) and Linnea Bergen (car 6). Considering that his job is to maintain the hardware and software in GAStech, we speculate that his overtime work is related with system maintenance.

In the further investigation, we find out more clues about the huge transaction in “Frydos Autosupply”. When all of his trajectories on Jan 13th are drawn on the map, shown in Fig.5 (c), we notice that all of his movement paths are far from the auto supply store. As we known, $10,000 could support associated tools and equipment for a crime, such as hiring a truck for transporting abductees. Therefore, we wonder whether his credit card was abused.

# 5.3 The association analysis of group activities

A [carefully](http://cn.bing.com/dict/search?q=carefully&FORM=BDVSP6)-executed plan and a close cooperative team are the essential features of kidnapping, so it is important to identify group activities and analyze their relationships.

An interesting group visit is found out on Jan 10th. Most of the staffs in engineering department and IT group visited a same destination near L.Azada’s house around 19:00 (see purple bars in Fig.6 (a) and purple trajectories in Fig.6 (b)) and went back home one after another before 24:00. As Jan 10th is Friday, we speculate this might be a Friday party, like casino night or karaoke night.

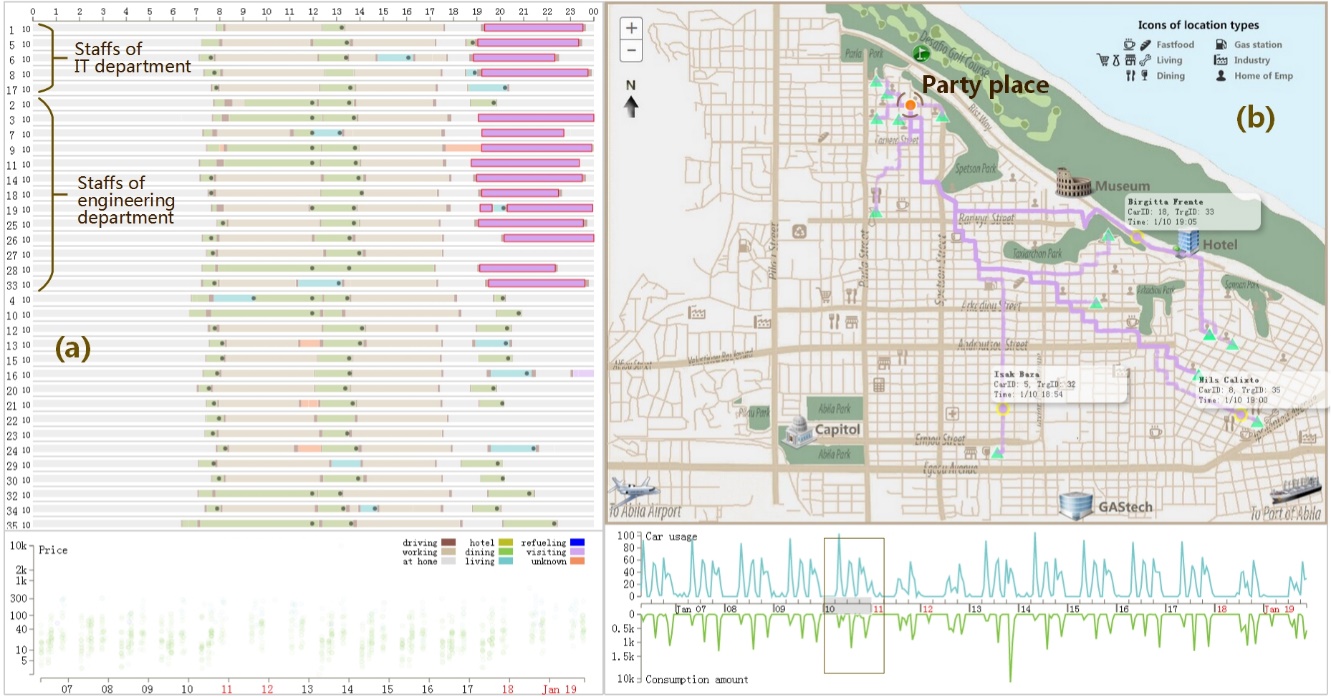


Fig.6 A Friday party on the evening of Jan 10. (a) The visiting activities of employees in IT department and engineering department (purple). (b) The destinations of these visiting activities are the same which is near L.Azada’s house.

The dangerous group activities involved four employees (I.Vann, Bodrogi, H.Osvaldo, M.Mies) in the security department. Fig.7 (b) shows the trajectories from midnight on Jan 6 to early morning on Jan 7th. From it we can detect that I. Vann went from his house to the place near CIO’s home at 23:00 on Jan 6th first, and then Bodrogi arrived at the same position at 03:00 on Jan 7th. On the morning of Jan 7th, they went to GASTech directly instead of returning home, which meant they had stayed near CIO’s house for the whole night. Similarly, the other three executive officers’ houses were visited by these four employees in three late nights (shown in Fig.7 (c)). For the job duties of I.Vann and Bodrogi are the management of surveillance equipment and alarm devices, we highly suspect that these four employees are related to some mysterious monitoring events on executives of GAStech.

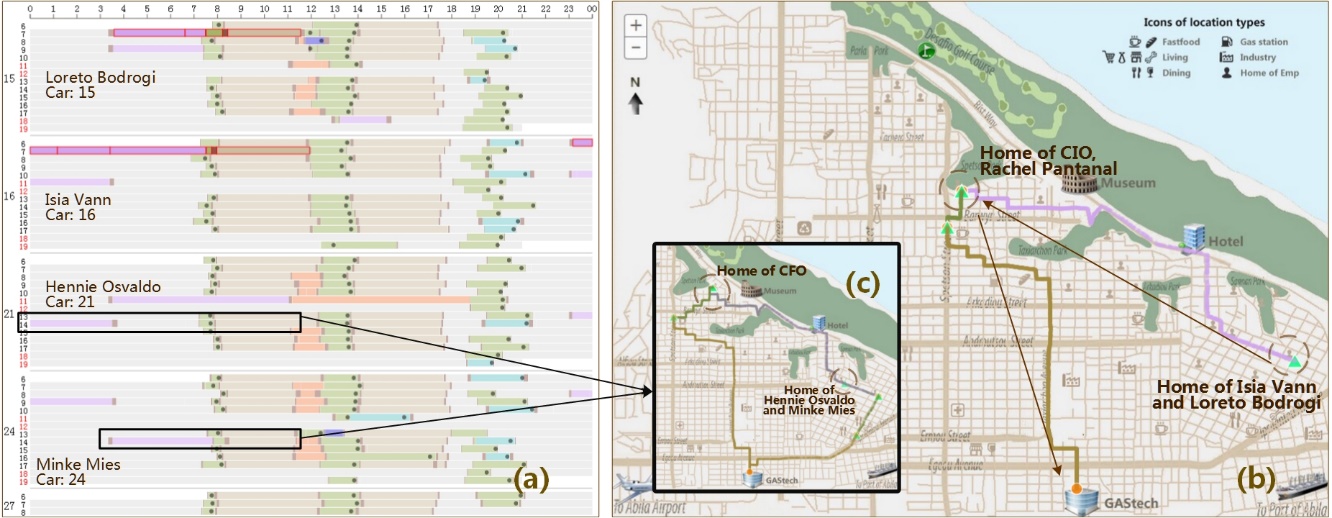


Fig.7 Monitored executive officers’ houses at midnight. (a) The abnormal visiting activities of four suspects at midnight (I.Vann, Bodrogi, H.Osvaldo and M.Mies). (b) I.Vann and Bodrogi monitored CIO’s house from 1/6 23:00 to 1/7 6:00. (c) H.Osvaldo and M.Mies monitored CFO’s house from 1/13 23:00 to 1/14 3:00.

In the further analysis, other dangerous group activities are found out. Shown in Fig.8, the three suspects in the monitoring events (Bodrogi, H.Osvaldo, M.Mies) and a new suspect (Inga Ferro, Female, Security, car 13) repeatedly visited five suspicious places about 11:00 without corresponding transaction records. Because these five places were never accessed by other employees, we strongly speculate that these places are related with masterminds of kidnapping, drug gangs or even illegal arms.

To sum up, the six suspects (I.Vann, Bodrogi, H.Osvaldo, M.Mies, I.Ferro and L.Alcazar) and four potential abductees (four executives of GAStech) are initially identified, and then the more targeted field investigations and face to face inquiries can be launched by the law enforcement.

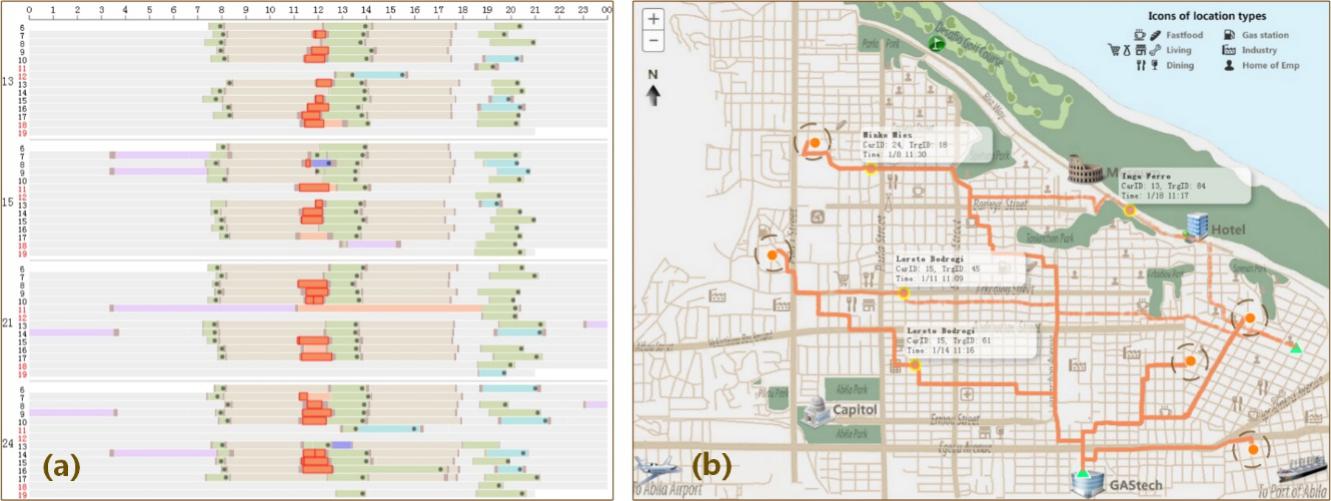


Fig.8 Five suspicious places visited by four suspects. (a) The abnormal activities of four suspects (Bodrogi, H.Osvaldo, M.Mies and Inga Ferro). (b) The positions of five suspicious places.

# 6 Discussion and future work

The system provided in this paper has obtained good result in solving the kidnapping case in MC2 2014. In this case, we listed the detailed information of the suspected persons, abnormal events and involved locations by priority. Compared with the ground truth, we found all the important places, five kidnappers and four abductees except for an abductee (Rachel Pantanal (CIO Assistant)), whose clues are not provided in MC2. However, there are still some limitations in this work which are considered as our further work.

The datasets provided in MC2 2014 can’t be considered as real big data. The car-tracking data is generated by 34 GPS-enabled cars for two weeks, including 680,000 GPS records. With the data preprocessing, about 6000 activity records are obtained for analyzing suspects’ activities. The efficiency of this system is quite good with the datasets after preprocessing. But when the data volume increases, the interactive efficiency, processing capacity and scalability may face some challenges. Consequently, it is necessary to improve the capacities of processing larger-scale crime-related datasets.

How to improve the usability of our system to support more real-world cases is an important question to consider. Databases similar to the car-tracking data in this paper should be well compatible in this system. The trajectory segment and trajectory painting in this system may not work well on sparse trajectory data such as the movement data of the mobile phone users recorded in theGSM cell. This problem may be solved by the integration of the trajectory aggregation techniques and the dynamic graph visualization techniques to explore patterns and study pattern connections. The transaction data as supplementary information in this paper is a typical data of recording human activities with temporal and spatial information. Thus, similar data such as the geotagged social events and the swiping card data in public transportation can also be compatible in this system. Since there are many different kinds of heterogeneous data in criminal investigation, such as text data (e.g. E-mail and news reports) and multimedia data (e.g. images and videos), the further researches should focus on how to improve the compatibility among different data and improve the processing [capacity](javascript:void(0);) of more heterogeneous data, to assist the trajectory data analysis.

In this article, the classification of the semantics is not comprehensive enough. It is designed according to the city and the company in which the kidnapping occurred. So this system can support the investigation of similar crimes occurred in enterprises and cities. However, the semantic types may [vary](http://cn.bing.com/dict/search?q=vary&FORM=BDVSP6) [wit](http://cn.bing.com/dict/search?q=with&FORM=BDVSP6)h different criminal backgrounds, such as the criminal site might be a farm or the border; the involved characters can be women or children. For the better application in actual criminal investigations, the semantic classification should be more integrated and reasonable; the match between the semantics and activities should be more intelligent.

The analysis process of this system starts from the time span filtering, then turns to the filtering of the characters and semantic activities, the last part is the exploration of geographical positions. This analysis process should be more flexible and operable. For example, the analysis can starts from the interested geographical areas. The abstract representations of the trajectories such as heat map are not adopted in this system since the analysis scheme of our tool is focused on the trajectory details. The abstract representations should be adopted if the analysis starts from the geographic areas.

# Conclusions

This paper proposed a visual analysis approach for the police or investigators to achieve a semantic enhanced analysis on trajectory data when detecting the kidnapping in MC2 2014. This paper first introduced the background, and analyzed the investigation requirements for the law enforcement. Secondly, in the part of data preprocessing, to solve the problem of the data [uncertainty](javascript:void(0);) and enrich the trajectory with semantic information, four steps were implemented: trajectory segment, location classification, geographic targeting and semantic assignment. Then, an interactive visual analysis system consisting of four views was presented to achieve an intuitive and informative trajectory analysis from the perspectives of time, space, person and semantics. In the case study, three cases were demonstrated to illustrate how this visualization system can help investigators obtain crime clues, lock the suspects, and gain evidence. In the last part, we discussed the limitations of our work and the future research on this subject.

# Acknowledgements

The authors wish to thank the anonymous reviewers for their comments. The authors would also like to thank the data providers, IEEE VAST Challenge. This work is supported by the National Natural Science Foundation of China under Grant Nos. 61103108 and 61402540, National Science & Technology Pillar Program of China under Grant Nos. 2012BAH08B01.

# References

Carrier B, Spafford E H (2003) Getting physical with the digital investigation process[J]. International Journal of digital evidence, 2(2), 1-20.

Casey E (2011) Digital evidence and computer crime: forensic science, computers and the internet[M]. Academic press.

Lang A, Bashir M, Campbell R, DeStefano L (2014) Developing a new digital forensics curriculum[J]. Digital Investigation, 11, S76-S84.

Beebe N L, Clark J G. A hierarchical, objectives-based framework for the digital investigations process[J]. Digital Investigation, 2005, 2(2): 147-167

Reith M, Carr C, Gunsch G (2002) An examination of digital forensic models[J]. International Journal of Digital Evidence, 1(3), 1-12.

Guo H, Wang Z, Yu B, Zhao H, Yuan X (2011) TripVista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection[C]. In: Pacific Visualization Symposium (PacificVis),IEEE, pp 163-170.

Liao Z, Yu Y, Chen B (2010) Anomaly detection in GPS data based on visual analytics[C]. In: Visual Analytics Science and Technology (VAST),2010 IEEE Symposium, pp 51-58.

Lee J G, Han J, Li X (2008) Trajectory outlier detection: A partition-and-detect framework[C]. In: Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference, pp. 140-149.

Aravecchia M, Calderara S, Chiossi S, Cucchiara R (2010) A videosurveillance data browsing software architecture for forensics: from trajectories similarities to video fragments[C]. In: Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence. ACM, pp 37-42.

Noblett M G, Pollitt M M, Presley L A (2000) Recovering and examining computer forensic evidence[J]. Forensic Science Communications, 2(4), 1-13.

Pollitt M (2010) A history of digital forensics[M]. In: Advances in Digital Forensics VI.Springer Berlin Heidelberg, pp 3-15.

Pasquale L, Yu Y, Salehie M, Cavallaro L, Tun T T, Nuseibeh B (2013) Requirements-driven adaptive digital forensics[C]. In: Requirements Engineering Conference (RE), 2013 21st IEEE International. IEEE, pp 340-341.

Guidance Software.EnCase Forensics - Computer Forensics Data Col-lection for Digital Evidence Examiners[EB/OL]. <http://www.guidancesoftware.com/encase-forensic.htm>, Accessed on Oct 2014.

Huebner E, Bem D, Bem O (2007) Computer forensics–past, present and future[J]. Information security Technical report, 8(2), 32-46.

Chainey S, Ratcliffe J (2005) GIS and crime mapping[M]. John Wiley & Sons.

X-Ways Software for Forensics, Data Recovery and ITSecurity X-Ways Software Technology AG [EB/OL]. <http://www.winhex.com/> Accessed on Oct 2014.

Mburu L, Helbich M (2014) Evaluating the Accuracy and Effectiveness of Criminal Geographic Profiling Methods: The Case of Dandora, Kenya[J]. The Professional Geographer.

Heim K (2014) Visualization and Modeling for Crime Data Indexed by Road Segments[D]. George Mason University.

Malik A, Maciejewski R, Collins T F, Ebert D S (2010) Visual analytics law enforcement toolkit. In: Technologies for Homeland Security (HST). 2010 IEEE International Conference ,pp 222-228.

Shiravi H, Shiravi A, Ghorbani A A (2012) A survey of visualization systems for network security[J]. Visualization and Computer Graphics, IEEE Transactions on, 18(8), 1313-1329.

Choi H, Lee H, Kim H (2009) Fast detection and visualization of network attacks on parallel coordinates[J]. computers & security, 28(5), 276-288.

Zhao Y, Liang X, Fan X, Wang Y, Yang M, Zhou F (2014) MVSec: multi-perspective and deductive visual analytics on heterogeneous network security data[J]. Journal of Visualization, 17(3), 181-196.

Andrienko G, Andrienko N (2008) Spatio-temporal aggregation for visual analysis of movements[C]. In: Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium, pp 51-58.

Scheepens R, Willems N, van de Wetering H, van Wijk J J (2011) Interactive visualization of multivariate trajectory data with density maps[C]. In: Pacific Visualization Symposium (PacificVis). 2011 IEEE, pp 147-154.

Tominski C, Schumann H, Andrienko G, Andrienko N (2012) Stacking-based visualization of trajectory attribute data[J]. Visualization and Computer Graphics, IEEE Transactions on, 18(12), 2565-2574.

Wang Z, Lu M, Yuan X, Zhang J, Wetering H V D (2013) Visual traffic jam analysis based on trajectory data[J]. Visualization and Computer Graphics, IEEE Transactions on, 19(12), 2159-2168.

Song X, Zhang Q, Sekimoto Y, Horanont T, Ueyama S, Shibasaki R (2013) Modeling and probabilistic reasoning of population evacuation during large-scale disaster[C]. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining.ACM, pp 1231-1239.

Song X, Zhang Q, Sekimoto Y, Shibasaki R (2014) Prediction of human emergency behavior and their mobility following large-scale disaster[C]. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp 5-14.

Krüger R, Thom D, Wörner M, Bosch H, Ertl T (2013) TrajectoryLenses–A Set‐based Filtering and Exploration Technique for Long‐term Trajectory Data[C]. In: Computer Graphics Forum (Vol. 32, No. 3pt4). Blackwell Publishing Ltd, pp 451-460.

Andrienko N, Andrienko G, Fuchs G (2013) Towards Privacy-Preserving Semantic Mobility Analysis[C]. In: EuroVis Workshop on Visual Analytics.The Eurographics Association, pp 19-23.

Krüger R, Thom D, Ertl T (2014) Visual analysis of movement behavior using web data for context enrichment. In: Pacific Visualization Symposium (PacificVis). IEEE, pp 193-200.

VAST Challenge Homepage [EB/OL]. <http://vacommunity.org/VAST+Challenge+2014> Accessed on Oct 2014.

Garfinkel, Simson L (2010) Digital forensics research: The next 10 years[J]. Digital Investigation, 7, S64-S73.

Schreck T, Bernard J, Von Landesberger T, Kohlhammer J (2009) Visual cluster analysis of trajectory data with interactive kohonen maps[J]. Information Visualization, 8(1), 14-29.

Boyandin I, Bertini E, Lalanne D (2010) Using flow maps to explore migrations over time[C]. In: Geospatial Visual Analytics Workshop in conjunction with The 13th AGILE International Conference on Geographic Information Science (Vol. 2, No. 3).

Kapler T, Wright W (2005) GeoTime information visualization[J]. Information Visualization, 4(2), 136-146.

Pirolli Peter, Card S (2005) The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis[C]. In: Proceedings of International Conference on Intelligence Analysis (Vol. 5). pp 2-4.

Ester M, Kriegel H P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Kdd (Vol. 96). pp 226-231.