T-Watcher: A New Visual Analytic System for Effective Traffic Surveillance

Jiansu Pu[1], Siyuan Liu[2], Ye Ding[1], Huamin Qu[1], and Lionel Ni[1]

[1]The Hong Kong University of Science and Technology, [2]Carnegie Mellon University

[1] {jspu,valency,huamin,ni}@cse.ust.hk, [2] syliu@andrew.cmu.edu

Abstract—Nowadays, big cities are suffering from severe traffic congestion as a result of the continuing increase in vehicles. Taxis equipped with GPS can be viewed as sensors of the traffic situation in city. However, trajectory data generated by taxi's GPS traces are often high-dimensional and contain large spatial and temporal attributes, which pose challenges for analysts. In this paper, based on taxi trajectory data, we present an interactive visual analytics system, T-Watcher, for monitoring and analyzing complex traffic situations in big cities. Users are able to use a carefully designed interface to monitor and inspect data interactively from three levels (region, road and vehicle views). We develop a visualization method to monitor and analyze traffic patterns for abnormal behaviors detection. In the region view of our system, global temporal changes in spatial evolution will be presented to users and can be interactively explored. The road view shows temporal changes to the traffic situations of significant segments of roads. The vehicle view uses a novel visualization method to track individual vehicles. Furthermore, the three views integrate important statistical and historical information related to traffic, which illustrate temporal changes of the traffic. We find that this design can help users explore historical information while monitoring traffic. We test our system on a real-life vehicle dataset collected from thousands of taxis and obtained some interesting findings. The experimental results confirm the effectiveness and efficiency of the proposed visual detection method. The analysis of the results also shows that our system is capable of effectively monitoring traffic and detecting abnormal traffic patterns.

I. INTRODUCTION

Nowadays vehicle movement patterns can be captured by trajectory data. Movement patterns are important for traffic analysts to understand the behaviors of moving objects especially in transportation management. Monitoring and analyzing trajectory data could be used to reason about mobility and to support experts in traffic analysis with reliable information. For example experts in transportation department can figure out why congestions happen more frequently nowadays and find an effective way to ease the traffic load in the modern cities. One effective way to understand traffic situation and vehicle status on road networks is monitoring and analyzing the trajectory data generated from taxis equipped with GPS. Taxis can be utilized as mobile sensors to constantly probe the city's traffic flows and assess the citywide traffic situation [16], [17]. We can assume the knowledge about taxi diaspora and behavior extracted from these GPS data is also valuable for traffic analysts, so as to analyze congestions, make route suggestions in the rush hour, and detect changes in the movement behavior as they occur.

However, trajectory data analysis now faces some techni-

cal challenges. As trajectory data contain both spatial and temporal attributes and are often huge in size and high in dimensionality. Thus large-scale trajectory data analysis is a very challenging task. For example daytime behaviors may differ from nighttime behaviors, while working days differ from weekends and holidays. Therefore a human analyst's sense of space and place is required, which is hard for a machine to achieve. It is important to keep humans in the analysis loop to utilize their sense of space and place, their tacit knowledge of inherent properties and relationships, and space-related experiences. Visual analysis of trajectory data shows great potential as they can intuitively present trajectory data and provide rich interactions, allowing users to explore the data. Historical data are also very helpful to analysts since humans need hints such as visual clues or visual displays of past information. We should provide the traffic analysts a new visual analysis system to not only intuitively visualize large-scale data, but also show the complex features and hidden patterns.

To tackle the challenges and assist in the understanding of trajectory data to improve traffic analysis we develop an interactive visual analytics system, T-Watcher, for monitoring and analyzing complex traffic situations in big cities via taxi trajectory data for Regions, Roads, and Vehicles, as illustrated in Fig.1. The developed visualization system enables a user to investigate trajectories at three different levels including in a region, on a road, or individual vehicles. The sophisticated fingerprinting method can well explore the spatial, temporal and multi-dimensional perspectives of the trajectory. It can provide more statistical information and transform the numerical knowledge to visual cues like shape, color, and size and so on. So users can easily analyze any changes in the spatial situation (inside the road network) over time, or to analyze the temporal changes in traffic situations over a road segment, or to track individual vehicles' instant status with historical data. In our system, users can perceive the correlations among different attributes and filter out noise and irrelevant trajectories for further investigation of interesting cases. Analysts can interactively and progressively refine the settings to improve the results.

We use three case studies to evaluate our system and demonstrate our fingerprint design on real-world taxi GPS data sets from 15, 000 taxis running for 92 days over a non-continuous eight months in a Chinese city with a population of over 10 million. Experiments show that our system is capable of effectively finding regular patterns and anomalies in traffic flows. In
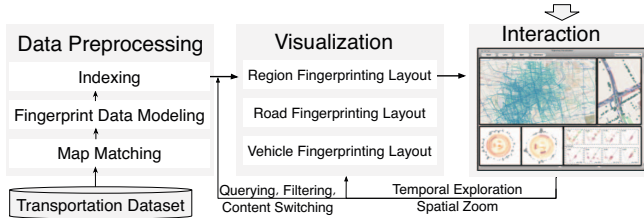
IEEE
computer
society

Fig. 1. The system architecture. The T-Watcher architecture consists of three primary components: (1) a data preprocessing module, (2) a visualization rendering module, and (3) a user interaction module.

summary, we have made the following contributions. (1) We have developed a system to study large-scale transportation data, integrating visualization and data analytics methods, and propose techniques to improve efficiency and scalability. (2) We report our experience and observations in building T-Watcher. (3) We demonstrate and test our system using real-life data sets for real-world applications.

II. RELATED WORK

Geo-visualization provides interactive visual tools for exploration and analysis of data with geographical information. This is a broad and extensively studied field. Due to limited space, we only summarize a few representative papers. Geographic visualization involves the interactive exploration of geographically-referenced information graphics to visualize data involving geographic features. A number of techniques and methods adapted from cartography and scientific visualization are studied and applied in geographic visualization. Mehler et al. [18] represented news sources as datamaps to show the geographic popularity of an entity, and any possible geographic bias. Wood et al. [21] discussed geo-visualization mash-up techniques including tag clouds, tag maps, data dials, and multi-scale density surfaces for exploratory visual analysis of large spatial-temporal datasets. Fisher presented hotmap [9] to represent aggregate activity and draw users' attention to the map. One geographic visualization challenge lies in visual thinking and user interaction, such as how human vision perceives maps and images and how it finds patterns. Chang et al. [7] presented legible cities to display large collections of data for urban context with different levels of abstractions. Some key issues and typical visualization approaches for geo-visualization of dynamics, movement and change are summarized in [4]. The above approaches keep the spatial layout as the indicating reference, whereas several layout-independent methods are proposed for geographic visualization. Growth Ring Maps [5] is proposed to represent spatial-temporal data by plotting a number of non-overlapping pixels. In comparison, our work is designed and proposed for large-scale trajectory data in an urban size, including both trajectory and road networks data. We proposed to provide interactive geographic visualization for similarities exploration and pattern extraction of interest in spatial temporal data. Crnovrsanin et al. [8] introduced a proximity-based visualization technique to discover the human behavior patterns from movement data. This abstract space is obtained by considering proximity data, which is computed as distances between entities and some important locations. The raw position is transformed into an abstract space such that the geographical information is transformed into meaningful multivariate data. Andrienko et al. [4] summarized the approaches in visualizing movement data. Characteristics of movement data and methods to present dynamics, movement, and change are discussed. In [2] they also surveyed the existing approaches to aggregation of movement data and visual exploration of the aggregates. The data were aggregated by predefined areas. It also presented specific methods and demonstrated the use of various visualization and interaction techniques. These authors also defined aggregation methods suitable for movement data and proposed interaction techniques to represent results of aggregations, enabling comprehensive exploration of the data in [1]. GeoTime [12] displays the 2D path in a 3D space to provide a detailed view of the geographical and temporal changes in movement data. Willems et al. [20] visualized vessel movements as well as the vessel density along traces by convolving trajectories with a kernel moving with the speed of the vessel along the path. Guo et al. [10] presented a trajectory visualization tool that focuses on visualizing traffic behavior at one road intersection. The spatial and temporal views are separated and the user can interactively explore the moving patterns of the trajectories.Microsoft T-drive [22] makes recommendations of the fastest paths taken by taxi drivers. The path is computed based on historical trajectories data. To select a few interesting trajectories from a large number, Hurter et al. [11] proposed a brush-pick-drop interaction scheme to visualize aircraft trajectories. The system supports the display of multiple trails and the altitude of each aircraft. Their methods are focused on 2D trajectory data exploration with limited perspectives provided. We tried to provide more comprehensive perspectives for trajectory data exploration in our system. In our work, we not only visualize large-scale spatial temporal trajectory data sets, but also embed traffic analysis results to digital maps. We apply a multidisciplinary approach to develop a framework for the analysis of massive movement data taking advantage of a synergy of computational, database, and visual techniques. We introduce our framework and demonstrate its effectiveness by examples.

III. DATA AND SYSTEM OVERVIEW

Dataset The data used in this study is the taxi trajectory data collected from GPS in Shanghai, China. Our data record the trace of around 7,700 taxis and the sampling rate of our trace data is various from 20 seconds to several minutes. Each GPS record contains car ID, the latitude and longitude of the taxi, the date, the time of the day in seconds, the taxi's status (loaded / vacant) and the speed and the direction of the taxi. In this work, we adopted a Weighting-based map matching algorithm and an Interpolation algorithm to calibrate the erroneous and low-sampling-rate vehicle GPS trajectory data set. The details are available in our previous work [14] and [15]. In our proposed algorithms, we first integrated the vehicle GPS sampling data and digital road
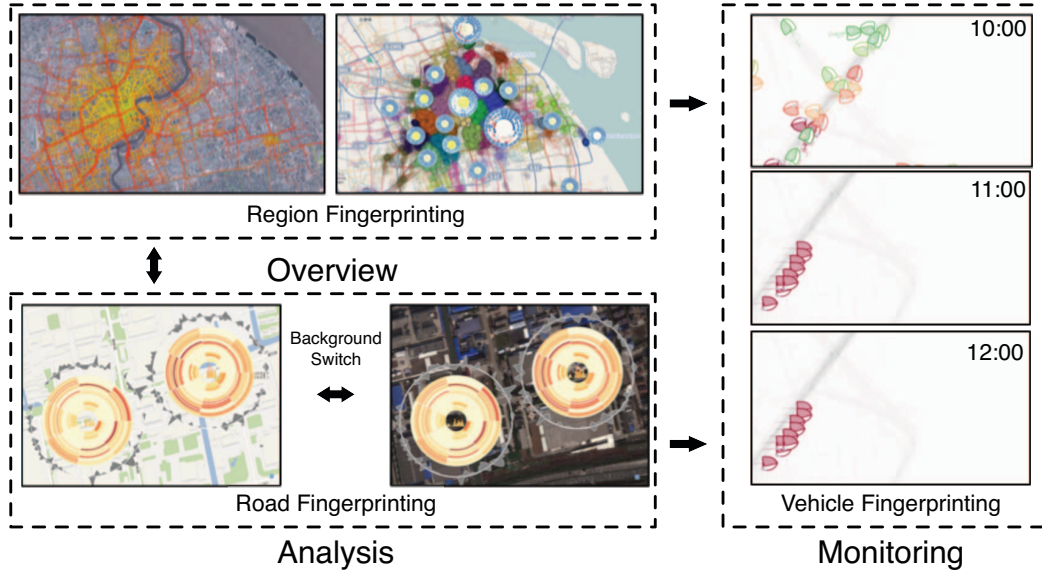
Fig. 2. The system's work flow. The system uses preprocessed vehicle GPS data as input and has three major components: the region fingerprint, the road fingerprint, and the vehicle fingerprint. The region fingerprint displays the whole spatial temporal distribution and uses the ring-map-based radial layout design to discover historical data. The road fingerprint can further analyze interesting locations selected from the region fingerprint. The vehicle fingerprint can help explore historical statistical information while monitoring the real-time situation.

networks data, to identify the road where a vehicle traveled and vehicle locations on that road. The statistical information for each road segment is computed as a preprocessing.

System Our system is implemented by using the "DaVincci" code package [6], an icon-based cluster visualization. The system has three major components: the region fingerprint, the road fingerprint, and the vehicle fingerprint.The whole system is designed to show the spatial temporal changes by presenting the instant values and historical data showing the evolution over a long time period. Fig.2 shows the flowchart of our system. Users first start from the region fingerprint. After the traffic data are loaded to our system to be analyzed, an overview of the traffic flow and the statistical information are displayed. Users are free to explore any interesting area and check any generated fingerprints for details. After that, users can select some interesting road segments for further investigation. When an interesting region is selected, the road fingerprints represent associated information such as the speed, time, and number of passing taxis. If a road segment in the region is interesting, users can choose to visualize the vehicle fingerprints for instant values in order to analyze any passing taxis' spatial and temporal attributes. All views support user interactions for interactive exploration.

Fingerprint Method We propose a novel trajectory fingerprinting method to discover essential characteristics, "fingerprints", by visually exploiting the multidimensional features in spatial temporal means which is proposed in our previous work [19]. Our fingerprinting method has the following benefits: (a) it leads to spatial temporal data feature extraction; (b) it provides a novel visual structure to answer queries by flexible and dynamic attributes combination; and (c) it is fast and scalable, easy to compare. The sophisticate

visual form is designed to extract spatial temporal features with multi-dimensions. And it can well explore the temporal related patterns including periodical patterns and frequent patterns. The "fingerprints" concept used here is tried to extract properties of good features by visually/visual analytics to compare trajectories. Good "fingerprints" also can help with anomalies detection and answering similarities queries.

IV. DESIGN AND IMPLEMENTATION

A. Region Fingerprint

The most natural way to represent the correlation for each region is to use a geographical map with visual items displaying the statistical information of each region. We consequently introduce a visual structure called the region fingerprint based on a ring-map design for interesting region exploration, which provides different attribute distribution using heat map layers as a background display. We want to fingerprint the hot spots of a city with a radial layout design to explore the related traffic data. Therefore in regional display, we apply our density fingerprinting method proposed in [19] to the traffic monitoring problem. The aggregation of taxis' spatial, temporal, or multidimensional information can be used to compute the hotness of each region and then a heat map is used to present the hotness of different regions. Then we need to display the data distribution of passing taxis over a selected 24 hour time period (week/month). The hourly distribution will be colored according to the defined color map and can reveal selected attribuer's correlation with the traffic. Hence we can analyze the local temporal attributes' changes and discover the evolution throughout each spatial region by comparing different fingerprints. In order to provide information
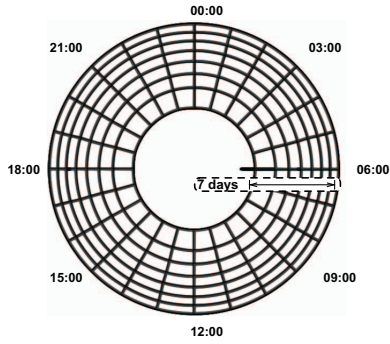
(a) Fingerprint Data Layout
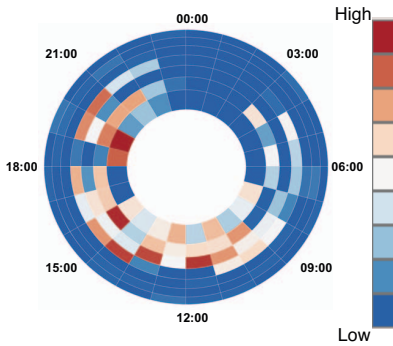
(b) Density Fingerprinting proposed in [19]

Fig. 3. The Region Fingerprint View Design. It uses a ring-map-based radial layout design to help explore historical statistical information. It shows the encoding scheme, which uses the typical heat map color, set represent density and the ring sectors around the fingerprint show the value of different time periods of the day.

at different levels of detail, we design a visualization of a fingerprint that adopts the radial layout design to display the 24-hour overview distribution for different regions of the city over a selected time period. Fig.3(a) shows our design layout. The circle is to encode a day or monthly distribution which starts from the top and numbers represent time. The time is shown on a circular axis with each big sector around the circle representing a day, which has 24 cells encoding 24 hours. And fig.3(b) presents our density fingerprinting display with color representing distribution values.

Ring Each ring on the fingerprint represents one day.

Sector Each sector inside a ring represents one hour, with time increasing by hours as the angle increases, and time increasing by days as the radius increases. The time of the region's behavior is displayed on each fingerprint's ring circle like a clock to encode a 24-hour distribution.

Color The color of the bar chart is an intuitive design to show the speed scale from green (high) to red (low), which is based on the color coding of traffic light signaling.

Size The fingerprint's size shows the total number of taxis passing through the selected area. For each fingerprint, we adopt a distortion method to allocate more display space to the inner sectors in order to have a better presentation.

Interactions For effective traffic data exploration and histor-

ical knowledge discovery, the region fingerprint design aims to provide users with, besides a good overview, zoom and details-on-demand by implementing brushing, linked view and focus+context view.

Brushing In the top-level fingerprinting display, the user can simply click on one region fingerprint to select it for further analysis. The user can also select the fingerprints in an interesting area or region by brushing i.e. users can use a mouse to draw a rectangle to select all the nodes inside that rectangle. When a fingerprint is selected, a new display will appear to display the selected region fingerprint as the center with nearby area (generally be displayed as the heatmap but user can specific this) and adjacent fingerprints. The user can then further explore the region they are interested in by sending the selected data to the road fingerprinting design.

Linked View Our region fingerprint can provide two spatial levels of traffic fingerprinting details. Meanwhile in urban scale we provide two kinds of display, heatmap-style fingerprinting and trajectory-style fingerprinting, which can be displayed simultaneously. In addition, the region scale traffic fingerprinting display can be presented to users at the same time by clicking a trip or a hot spot (grids) in the urban traffic-fingerprinting map, and then the corresponding region scale fingerprinting visualization will be shown on the map.

Focus and Context For the trips or locations that are of interest to users, our design can adjust the layout to provide a focus+context view, which gives more space to the trips or locations of interest and suppresses the space of other trips or locations as context.

B. Road Fingerprint

In this subsection we describe the road fingerprint design that visualizes the selected road segment's temporally related attribute distribution and can simultaneously show the temporal changes and local temporal variables in spatial evolution. From the region fingerprinting, one key observation is that traffic data especially taxis data are inherently visual as they are spatially temporal around road networks. Therefore we can assume that the overall traffic situation is composed of each road segment's traffic data. If one place or region experiences traffic congestion or other traffic phenomena we can also assume this road segment to be very important for analysis. Therefore, we want to visualize traffic data on specific road segments to provide more details for analysts besides just the region fingerprinting results. We need to further analyze each road's historical information. Therefore it is desirable to show both geographical and statistical information in one design. For each road segment, we would like to display the distribution of taxis, the time, the average speed, the pick-up/drop-off passengers behavior changing or evolution and the geographical information of any road of interest. We propose a similar visual approach with the region fingerprint design but using a different visual encoding scheme and a new visual component called "similarity lens" including a "routine indicator" and "abnormal detector", at outer glyphs.

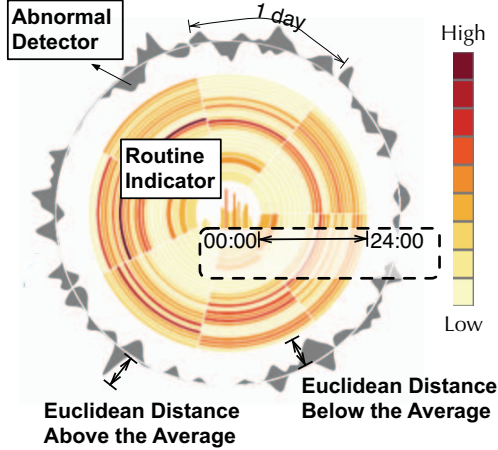Fig. 4 shows the road fingerprint. Our design displays the

Fig. 4. Road Fingerprint View Design. The road fingerprint displays the routine pattern in the center, surrounded by temporally distributed historical attributes. A new visual component, "similarity lens", including a "routine indicator" and "abnormal detector" is presented. A routine indicator is bar chart like glyph structure and each bar represents a time period (hour or day). The height encodes the value. An abnormal detector is the outer glyph lying on the furthest ring. It has a routine line which indicates the average value. Each big sector on the cycle represents a day that has a related mapping part on the glyph.

routine pattern inferred from statistical representation in the center with historical attribute information temporally distributed around different time granularities, either on a 24-hour scale or a 31-day scale based on users' choice. Here we define the routine pattern as the average value. As most of the statistical information is associated with time, we propose a similar design as [13] and use a radial layout to show the statistical information distributed over time. We use linked views here to further explore the correlation among different attributes of taxis i.e. speed and pick-up/drop-off behaviors. We use color to intuitively identify attribute values. The time of the taxis' behavior is displayed on the node's circle like a radial calendar to encode time distribution. It is different from the region fingerprinting view and each big sector on the cycle represents a day. The cells inside each sector plot one-hour value.

Sector Cells Each cell inside the big sector represents an hour. The color of the cell is intuitively designed to show the value.
Outer Glyph The part extruding from the fingerprint represents a related value larger than average. In contrast, the part extruding to the center of the cycle represents a value below average. Each big sector on the cycle represents a day that has a related mapping area on the glyph.
Inner Glyph A bar chart like structure is used to form a glyph to encode the routine data. Here we define "routine" as the average value. Each bar height represents the average value of that time (hour or a day) and the color scheme is the same as with the fingerprint.
A new visual component called "similarity lens" including a "routine indicator" and "abnormal detector" is designed to encode routine pattern information. The routine indicator is a bar chart like glyph structure. It takes a series bar chart in
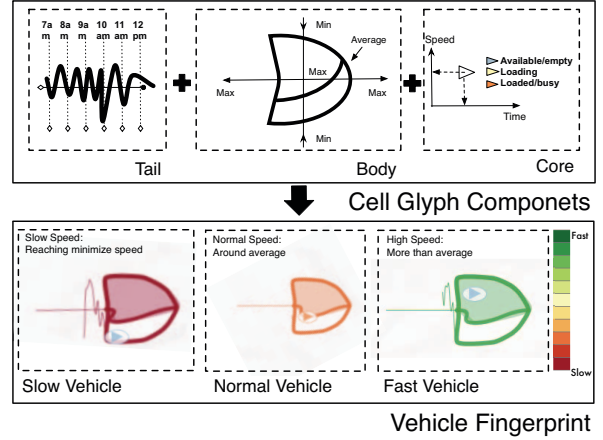

Fig. 5. Vehicle Fingerprint Design. Time stamp is encoded by the arc's horizontal position. The time starts near the tail and ends at the head. The arc color indicates taxi status. The tail encodes the taxi's speed changes in the historical data. One straight line indicates the average speed. The body shape and board color are used to encode instant speed values. Inside the body an inner coordinate system is designed to use the core's position in a vertical direction to identify whether the instant speed is above average, near maximum, or reaching the bottom.

different granularities of time, either on a 24-hour scale or a 31 day scale based on users' choice. In addition, the height and color both encode the attribute values. The abnormal detector is the outer glyph located at the furthest ring of the road fingerprint. It has a routine line that indicates the average value and each big sector with one day's 24 cell data plotted on it has a related part on the glyph. This part has 24 points on it. If the related time data recorded at 4*am*m is over the average, this part of the glyph will extrude outside the fingerprint at the 4*th* point. If it is below average, it will extrude inside the fingerprint at the 4*th* point.
Interactions Some user interactions are supported in our road fingerprint design to allow users to select interesting time periods for historical information to compare with real-time traffic data. There are two ways to explore the historical information of an interesting time period. Users can use a mouse to draw a rectangle on the time circle to brush the time periods. Then the related hour or day in the inner glyph and outer glyph will be highlighted.
Linked View The road fingerprint with different data attributes like speed, or vehicle density can be presented to users at the same time by clicking a road fingerprint in the map, then new views will appear with another fingerprinting visualization representing different attributes.
Union Selecting The two brushing interactions also support union. Users can add new time periods to previous selections.

C. Vehicle Fingerprint

In this subsection we present our novel cell-shaped-based layout that visualizes each taxi's historical information with its real-time traffic situation. To assist in understanding or future improvement of traffic situation, we developed Cell-glyph, a visual fingerprint for instant vehicle values and historical

data discovery, specifically GPS data from taxis. In previous sections, we demonstrate how our fingerprinting method can reveal historical information and discover related knowledge with historical data at different levels of detail. However, we still find that we need another visual structure to present each taxi's historical information with its instant data. We develop a novel visual structure, cell-shape-based layout, which comes from the biological concept of a cell shape. We use this new visual structure to encode the real-time traffic situation together with historical knowledge in one structure. We display the time, the instantaneous speed, the direction, and statistical information. Our novel visual structure takes the concept of a biological cell to encode instant changes with historical knowledge. Each cell unit is used to encode instant changes of one passing taxi in a selected region or road segment. The cell-glyph design promotes the exploration of the knowledge of an activity or behaviors of monitored taxis based on historical information. Here, our vehicle fingerprint has chosen the metaphor of biological cells, to graphically and dynamically convey historical information and speed changes that have happened in the past. As a result, its graphical style is unique, to say the least. In terms of visual representation, each "cell" in our system represents a particular taxi. Bigger cells represent taxis, which have recently recorded more information. Each glowing "particle" inside a cell/fingerprint represents a recent update from that respective taxi. The more a particle moves around, the more active or mobile it is.

Inner arc We use an arc to replace the cell core to encode the taxi's instant direction. And the time of the taxi's information is encoded by the inner arc's position. The time starts at the point near to the tail and ends at the head. The color of the arc represents status of the taxi, which green means available, yellow means loading on/off passengers, and red means passengers loaded.

Tail It encodes the taxi's speed changes in the history data. And the straight line indicates the average speed.

Cell Shape The width and the height are both used to encode the instant speed value and the color of the cell board is an intuitive design to show the speed. As stated before, red means low speed and green means high speed.

Inside Cell An inner coordinate system is designed to use the core's position in a vertical direction to identify whether the instant speed is above average, near maximum, or reaching the bottom. Also we use a curve inside the cell body to indicate the average value and we color any area that represents values larger than average. So if the arc lies in the colored area we intuitively know this taxi's current speed is higher than its average speed.

V. SYSTEM EVALUATION

Experiment Setup The experiments are conducted on an Intel(R) Core(TM) 2 2.13GHz PC with 1GB RAM and an NVIDIA Geforce 7900 GS GPU with 256MB RAM. We employed JDBC to connect to the database server and Java swing to develop visualization modules. The digital map is developed based on an open source map (OpenStreetMap,
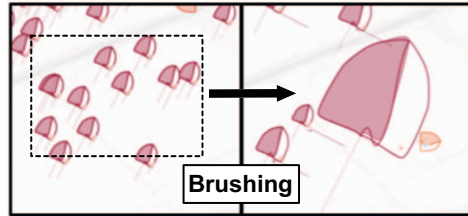

Fig. 6. Automatic Vehicle Fingerprint Grouping Approach. Users can brush vehicle fingerprints to get some aggregate fingerprints. The aggregate fingerprint is proportional in size to the number of vehicles it is grouped in, and the color of the board, the tail will be changed to display the average values of the group.

including bit maps and vectors to describe the objects in the digital map). After the map-matching algorithm sanitizes the data, we are able to get the valid taxi trajectory data. The statistical information for each road segment is then computed as preprocessing. The preprocessing dealing with the data of thousands of taxis in one month took about six hours. After that, the system supports interactive query. After the preprocessing, our system supports interactive real-time visual displays and user interactions. In this section, we report our findings using our system to analyze the taxi trajectory data collected from a non-continuous eight-month period of 92 days in total.

A. Region Fingerprint Results

Validation For testing our design, we used a dataset where certain spatial and temporal patterns were previously expected. The typical temporal patterns of traffic situations in a big city are well known. Thus, there are particular intervals in the mornings of working days, called "rush hour", when the major arterial roads are crowded with vehicles and movement is obstructed. Similar situations occur in the spatial locations. We can always expect some regions in the city to have more passengers looking for taxis and experienced taxi drivers prefer to adopt pathways to satisfy passenger requirements such as the trip time to the airport. We tried to detect these expected temporal patterns by comparing similar spatial situations with the help of our region fingerprint. We had discussed the hot spot exploration in our previous work [19]. But it is not too intuitively for analyzers to discover the correlation between taxi records with the temporal evolution. So we further improved our method on those findings and had a better explanation on these results. We can easily figure out hot spots in which higher percentages of customers are picked up/dropped off by taxis traveling around the urban area (Fig. 7). We first used a density-based partitioning algorithm K-mean for hot spot computing by rasterizing the background map into pixels with GPS values. We labeled each pixel in the map by assigning the vehicle values to them then the generated clusters naturally formed as regions. Clusters were colored as different hues to identify them. Then we computed the total number of customers being picked up/dropped off by taxis for each region. Then we used our region fingerprinting results to reveal the hotness of regions over the 2D map of the city. Fig. 7 proved the hot spots identified by region

fingerprints. We used filtering to display only top 5 hot regions in the week (See (a)(b)(c)(d)(e) in Fig. 7). It is obviously each generated fingerprint on the map is colored by a dense hue which means this spot had contributed more values to the traffic. And the detail situation like the daily or weekly display data distribution can be observed from the visual cues (color) on the fingerprints. For example we can find out the region (b) is at the airport so we can expect many travelers taking taxis there. Region (d) is at the main roads where a crossing bridge over river located nearby. So we can observer the fingerprint that most pick-up/drop-off behaviors happened in the day time. Users can use the partitioned map to choose some interesting regions to explore their details or change to the road fingerprint view for further analysis. The results clearly show that our region fingerprint design can predict spatial and temporal patterns.

Application In this part, we used the region fingerprint to explore a hot spot identified from the region fingerprinting results (Fig. 7). We wanted to figure out whether hot spots in a similar density have similar data distributions or not. In this case, we designed the fingerprint within one week for all 65,836 roads with all 3 attributes (vehicle id, average speed, and picking-up / dropping-off id). In order to explore these hot spots that are identified we draw the related region fingerprinting results (see Fig. 7). Taxi pick-up / drop-off fingerprinting results are a surprise since they prove the clusters in a similar dense varying in distribution. As shown in Fig.7, the advantages of our method are clearly revealed and show the results of hot spot exploration. We find the results are out of our expectations, especially when comparing the fingerprints in region (d) and (e). Even thought they are of similar density (this can be observed from the size of the fingerprints) the fingerprinting results revealed different behaviors. According to the colors of the rings, we find taxi pick-up / drop-off distribution varies in different regions over different time periods if the colors are not even, while in the center of the city they were almost the same when the colors are smooth. Another interesting finding is that our top 1 hot spot in region (a) which had most taxi pick-up / drop-off values than other regions in a rather small place which was far away from city center (almost reaching rural area). And from the Fig.7, we found the top spot in the map only occupied a small area only consisting two roads. It is clear that this hot spot must have some interesting patterns hidden behind the raw data. We can further check this spot in a more detail level by our road fingerprinting view.

An integrative approach was employed in [3] by combining self-organizing map (SOM) with a set of interactive visualization tools. They put both feature and index images separately into SOM matrix cells to give a combined representation of the spatial, temporal, and attributive (thematic) components of the data. Another data aggregation approach used the predefined areas in [2]. It applied pixel-based visualization to show the aggregated temporal changes to each grid in the Milan map and the spatial evolution of local temporal variables is clearly visible. In our system we can choose the attributes to be
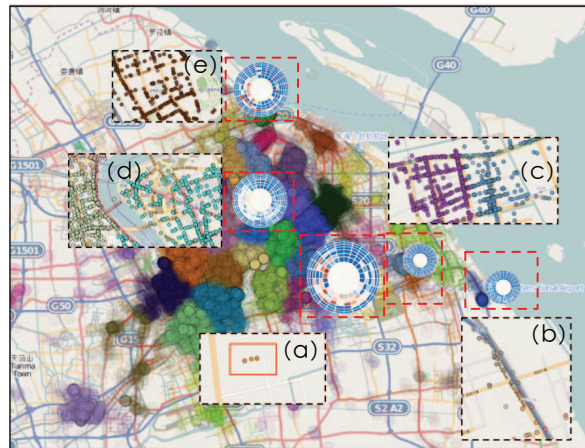

Fig. 7. Hot Spot Exploration by region fingerprint. We applied the region fingerprinting view on taxi pick-up/drop-off density to explore hot spots in the city. Fingerprints for top 5 hot spots are displayed.

presented in a more flexible way and utilize the spiral layout to better explore temporal patterns like periodic events.

B. Road Fingerprint Results

Validation An obvious spatial pattern that can be expected in the distribution of the local temporal variations is that the traffic on the major roads differs from that in the city center (revealed in Fig. 7). One can also expect a frequent or periodic pattern in the traffic data in some areas e.g. resident areas. To detect such patterns, we compared the temporal variations in each selected road segment with the help of a road fingerprint by using the same parameters in the previous experiment. As shown in Fig. 8 (c), we visualized the pick-up/drop-off hot spots from the heat map exploration result (right bottom, dark red color spot). We have observed that each of the two regions in this spot was occupied by one road, the road fingerprint located on the right-side showed a clear periodic pattern where the dense color formed a strip around noon. This pattern means that in this road the pick-up/drop-off behavior usually reaches its peak around the middle of the day and can be stable during the week.

Application We want to detect any possible anomaly patterns from the traffic data by using our fingerprint design to identify the details. Here we define anomaly patterns as anonymous errors, unusual phenomenons from normal distribution data. We started with our region fingerprint with density map (Fig. 7). Here we use speed as the variable visualized as a density style map to explore the patterns. Speed, viewed as a dynamic vehicle indicator is very unique, which gives us a multitude of other information, such as spatial information, temporal information or the behaviors of the vehicles. We checked abnormal spots discovered from the speed density map by using the road fingerprint for further investigation. Our road fingerprint design can help analysts to distinguish the details of abnormal patterns from routine behaviors. In this case, we found some interesting abnormal patterns inside road segments. We believe it is good for our road fingerprinting to

reveal such knowledge since we can compare them with the history data on a rather large scale say of a whole month to check whether it is an event or not. Fig. 8 shows the details of these findings, which formed a really strange pattern. We can clearly see in Fig.8 that no matter how the vehicle density changes, if the average speed is greater than zero, they appear the same hue and saturation of color, for example, the green in the figure. We look at the vehicle fingerprinting of this case and observed that the abnormal detector formed a smooth cirque (see fig. 8 (a)) without any roughness. Our abnormal detector provides an intuitive view on the data's Euclidean distance to average value, meaning that the average speed per hour that week on these two road segments is nearly the same. This may not be possible so we recheck the original data, but then found the records were mostly contributed by one single taxi. The records showed that the driver shifted several times in one hour between the two roads, therefore some peaks formed in the data since the records were affected by the rather stable average speed ($11km/h$). We also checked the driver's history data. His taxi's mobility was high and we picked a specific hour to check whether it appeared in two places simultaneously. Here we applied our vehicle fingerprinting technique to better reveal the result. We found he appeared in at least three different places, therefore it turns out that these peaks are definitely hidden errors of the data. The spiral layout can make data periodic trends easily apparent when the correct period is chosen. In [23] the authors emphasized activity and cyclic time as dominant issues in its representation and it was designed to explore patterns from the timeliness of movements, availability and events. Its visual component structure is similar to our road fingerprint design. However, our system view is focused on providing different hierarchical levels of temporal attribute exploration over a single road segment. It can also reveal a one-month pattern of the selected segment, and simultaneously provide visual analytics from multiple aspects of visualizations of spatial, temporal and multi-dimensional perspectives that are linked together.

C. Vehicle Fingerprint Results

Validation Nowadays the considerably increased number of automobiles in cities has exponentially congested transportation pathways. Heavy traffic may give rise to a significant safety risk and lead to the urgent requirement of advanced technologies for us to monitor and model real-time traffic data in today's transportation management. We also want to set up related experiments to show our approach can be helpful in traffic monitoring. We are interested in unusual traffic phenomena e.g. sudden high speed, exceeding the speed limit, longer stops, and the like. We found some extreme cases for our statistical info, like 89 pick-ups / drop-offs in one hour located in one small region, taxis traveling at an average of $244km/h$, and 202 taxis recorded on the same part of the same road in one hour. However, we found when we analyzed these specific situations we lost any real-time information to allow to get the latest information. The converse of the monitoring
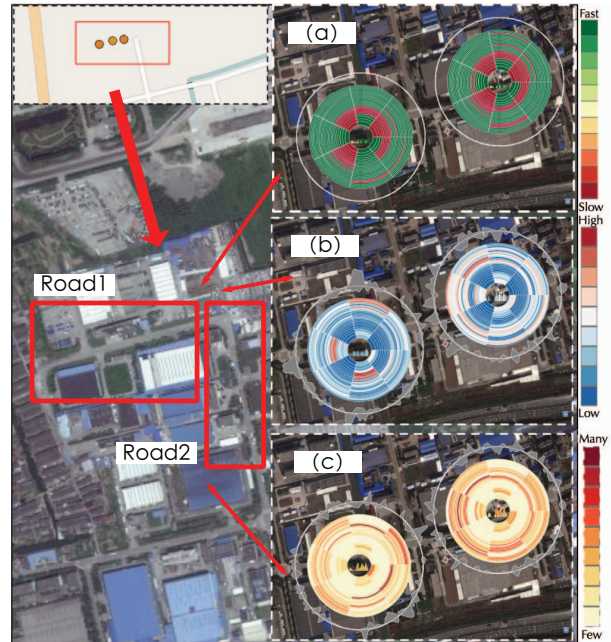

Fig. 8. Abnormal Data Detection and Identification. (a) Average speed distribution of two roads in one week. (b) Vehicle density distribution. (c) Pick-up/Drop-off distribution. We have identified an abnormal pattern where density changes did not affect the average speed on the road. It is rather stable when we check the abnormal detector forming a smooth cirque without any roughness. This means the speed per hour aligns very well with the average and it does not change greatly in the week.

situation is different, in that we lack historical information such as, average speed and past speed changes and so on. We applied our vehicle fingerprint to monitor the traffic. We have chosen three regions as shown in Fig. 9 to monitor: 1) the blue road with id 5588, where historical data reveals the number of vehicles as abnormal; 2) red road with id 7717, where the speed is quite high from the historic distribution, and 3) four roads 4693, 4694, 4695, and 4696, where the density of vehicles is abnormal and needs further exploration. For road 7717, we successfully detected the overdrive, and it is a sudden change via the tail as shown in Fig. 10. For road 5588, we also detected the abnormal alternation of the number of vehicles as shown in Fig. 11.

Application From taxi GPS records over the city, we can compute a number of instant characteristics. Instant characteristics include instant speed, direction, and the like. We believe the combination of displayed instant values with historical data can help further the knowledge in traffic analysis. From Fig. 10 and 11, we found that taxis with high mobility can be considered as sensors traveling around the city, which visually back up the assumptions in [22]. It is interesting and we can see that it visually supports the assumptions in [16]. They employed vehicles as sensors using their instant speed to sense the vicinity of vehicle congestion. They assumed the reported speeds are usually quite accurate because they are obtained directly from the speedometers installed on taxis and sudden changes in speed are rare. Our vehicle fingerprints can help to detect the traffic status by taking history data advantages into
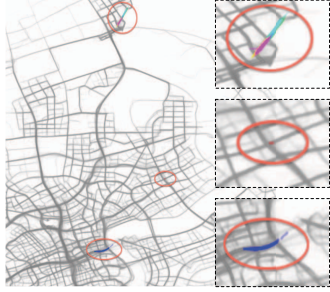
Fig. 9. The three regions we have chosen for traffic monitoring using our vehicle fingerprint design.
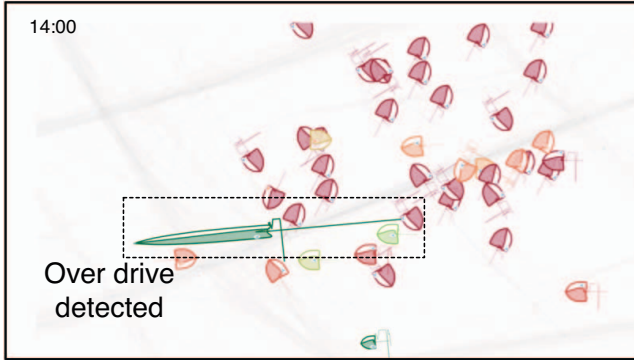

Fig. 10. Monitoring Result of Road 7717. It clears detect an over drive situation.

real-time monitoring. For roads 4693-4696 see Fig. 12(a), we have found many vacant taxis suddenly and rapidly emerge compared with routine situations. Meanwhile their tails on the fingerprint were curved indicating speed changes; hence we can say that they came from other locations. Similarly, we find that a lot of vacant taxis remain in the same location for hours, while other taxis passed by or left, but the majority taxis were static since their tails, where straight which indicates these taxis did not change speed for several hours. Based on this observation, we figured it to be a taxi terminal, which is true according to the corresponding historical data. We also checked the vehicle distributions of the selected week in the same place (see Fig. 12(b)). We found the same pattern constantly appeared around the observed terminal spot. These results showed the benefits of our design.

VI. DISCUSSIONS

Scale is the grand challenge of visualization. For the large scale trajectory data, visualization suffers from visual clutter and rendering efficiency problems. In our system, we propose

Fig. 11. Monitoring Result of Road 5588. It clearly detects sudden increases in taxi density at the usual time.

several novel visual encoding schemes for trajectory data analysis, and we also employ a set of visualization techniques. To address the scalability problem, our system follows the principle: overview first; zoom and filter; then details-on-demand. The system mainly consists of three components: region view, road view, and vehicle view. The region view provides a good overview of the different data attributes' distribution in a certain period of time. The overview describes an abstraction of the query result, providing the users an entrance to investigate the problems of their interest. The road view and vehicle view shows the query result based on the users' interaction in the overview. The road view reveals the correlations between the average value and history information, and then we can check the temporal distribution over the spatial to explore their spatial evolution. We introduce a novel visual structure called the cell-glyph for instant vehicle values fingerprinting, which can display real-time data with historical knowledge simultaneously. By interactively exploring the traffic data with the above visual displays, we are able to identify the correlations between the traffic and taxis' behaviors with clear patterns. We can further investigate the expected traffic patterns by applying the dataset to validate our design. We also discovered some hidden information and unexpected patterns that is, an anonymous error in the taxi trajectory data and unusual taxi terminal time with several attributes changing by using our fingerprinting design and we believe this can help data mining experts to explore traffic data. This method supplies a quick visualization process and user-friendly interface.

Since our system is developed based on a visual analytics idea, it may not be correct to use some of the traditional system evaluation metrics here. Our accuracy is relatively high since it utilizes a human analysts' intelligence to make the decisions. Analysts can refine or re-tune their results in an iteration way with the help of rich user interactions, so they can achieve satisfying results by progressively improving the parameters. The time cost for our data preprocessing is rather high but the query response time is acceptable with the filtering technique. We can apply our system to bigger datasets and achieve approximately the same results at a reasonable cost, since our system bottleneck is the data preprocessing and the visualization processing time cost is lower than we expected.

VII. CONCLUSIONS

In this paper, we have presented an interactive visual analytics system, T-Watcher, for monitoring and analyzing complex traffic situations in big cities via taxi trajectory data. Several new integrated traffic fingerprinting designs have been elaborated. We also designed a novel visual structure called cell-glyph to compare instantaneous situations with statistical information. Our system consists of three major modules (the region fingerprint, the road fingerprint, and the vehicle fingerprint). The region fingerprint allows users to investigate overall statistical information of important hot spots in the city and to suggest some interesting locations for further exploration. The road fingerprint displays both geographical and statistical

(a) Jan 23rd Monitoring Results
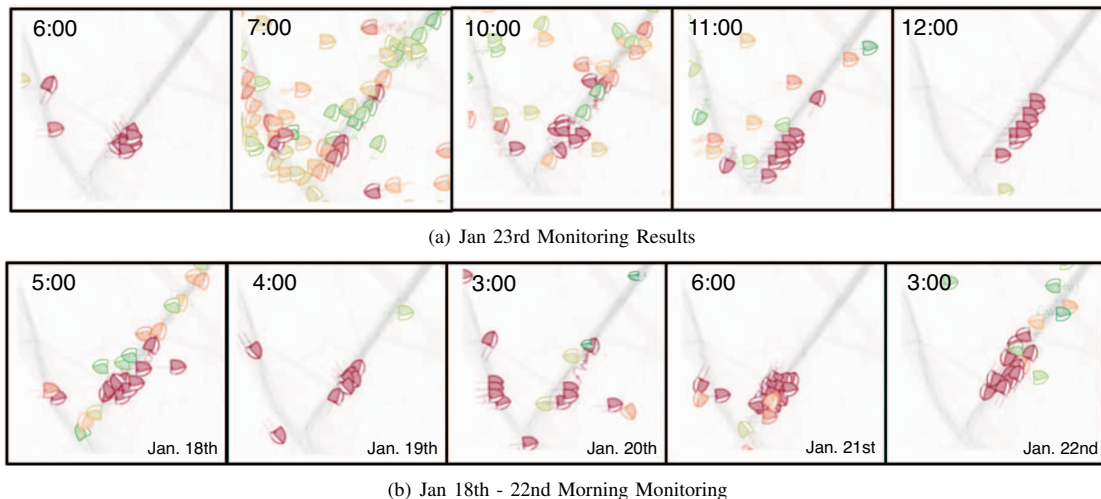

(b) Jan 18th - 22nd Morning Monitoring

Fig. 12. Monitoring Results of Road 4693 to 4696. It clearly reveals many vacant taxis coming in one afternoon and staying for hours. We guessed it might be a terminal for a group of taxi drivers. Historical data backed up our findings.

information with routine data. Finally, the vehicle fingerprint presents the real-time data with historic information, which greatly improves the monitoring. Because of the excessive amount of data, we applied a preprocessing method such as aggregation to reduce the scale of data to be visualized. We did not study the problem related scalability since we did not check our design with extremely large datasets. In future, we will investigate other effective methods to handle the scalability problem. We plan to add region separation and design a node projection algorithm to put nodes in better positions in the map.

REFERENCES

[1] G. Andrienko and N. Andrienko. Spatio-temporal aggregation for visual analysis of movements. In *IEEE Symposium on Visual Analytics Science and Technology, 2008. VAST '08*, pages 51 –58, October 2008.

[2] G. Andrienko and N. Andrienko. A general framework for using aggregation in visual exploration of movement data. *The Cartographic Journal*, 47(1):22–40, 2010.

[3] G. Andrienko, N. Andrienko, S. Bremm, T. Schreck, T. Von Landesberger, P. Bak, and D. A. Keim. Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns. *Computer Graphics Forum*, 29(3):913–922, 2010.

[4] G. Andrienko, N. Andrienko, J. Dykes, S. I. Fabrikant, and M. Wachowicz. Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research. *Information Visualization*, 7(3-4):173–180, 2008.

[5] P. Bak, F. Mansmann, H. Janetzko, and D. a. Keim. Spatiotemporal analysis of sensor logs using growth ring maps. *IEEE transactions on visualization and computer graphics*, 15(6):913–20, 2010.

[6] N. Cao, D. Gotz, J. Sun, and H. Qu. DICON: interactive visual analysis of multidimensional clusters. *IEEE transactions on visualization and computer graphics*, 17(12):2581–90, 2011.

[7] R. Chang, G. Wessel, R. Kosara, E. Sauda, and W. Ribarsky. Legible cities: focus-dependent multi-resolution visualization of urban relationships. *IEEE transactions on visualization and computer graphics*, 13(6):1169–75, 2007.

[8] T. Crnovrsanin, C. Muelder, C. Correa, and K.-L. Ma. Proximity-based visualization of movement trace data. pages 11–18, 2009.

[9] D. Fisher. Hotmap: looking at geographic attention. *IEEE transactions on visualization and computer graphics*, 13(6):1184–91, 2007.

[10] H. Guo, Z. Wang, B. Yu, H. Zhao, and X. Yuan. Tripvista : Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. *IEEE Symposium on Pacific Visualization, 2010. PacificVis'10*, pages 163–170, 2010.

[11] C. Hurter, B. Tissoires, and S. Conversy. FromDaDy: spreading aircraft trajectories across views to support iterative queries. *IEEE transactions on visualization and computer graphics*, 15(6):1017–24, 2009.

[12] T. Kapler and W. Wright. GeoTime Information Visualization. *Information Visualization*, pages 1–8, 2004.

[13] H. Liu, Y. Gao, L. Lu, S. Liu, H. Qu, and L. Ni. Visual analysis of route diversity. In *IEEE Conference on Visual Analytics Science and Technology, 2011. VAST'11*, pages 171–180, 2011.

[14] S. Liu, C. Liu, Q. Luo, L. M. Ni, and H. Qu. A visual analytics system for metropolitan transportation. In *Proceedings of the 19th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS 2011)*, pages 477–480, 2011.

[15] S. Liu, C. Liu, Q. Luo, L. M. Ni, and K. Ramayya. Calibrating large scale vehicle trajectory data. *Proceedings of the 13th IEEE International Conference on Mobile Data Management (IEEE MDM 2012)*, pages 222–231, July 2012.

[16] S. Liu, Y. Liu, L. Ni, J. Fan, and M. Li. Towards mobility-based clustering. In *Proc. of ACM SIGKDD*, pages 919–928, 2010.

[17] S. Liu, Y. Liu, L. Ni, M. Li, and J. Fan. Detecting crowdedness spot in city transportation. *IEEE Transactions on Vehicular Technology*, (99):1– 1.

[18] A. Mehler, Y. Bao, X. Li, Y. Wang, and S. Skiena. Spatial analysis of news sources. *IEEE transactions on visualization and computer graphics*, 12(5):765–71, 2006.

[19] J. Pu, S. Liu, H. Qu, and L. M. Ni. Visual fingerprinting: A new visual mining approach for large-scale spatio-temporal evolving data. *Proceedings of the 8th International Conference on Advanced Data Mining and Applications (ADMA 2012)*, December 2012.

[20] N. Willems, H. Van De Wetering, and J. J. Van Wijk. Visualization of vessel movements. *Computer Graphics Forum*, 28(3):959–966, 2009.

[21] J. Wood, J. Dykes, A. Slingsby, and K. Clarke. Interactive visual exploration of a large spatio-temporal dataset: reflections on a geovisualization mashup. *IEEE transactions on visualization and computer graphics*, 13(6):1176–83, 2007.

[22] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang. T-drive: driving directions based on taxi trajectories. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS 2010)*, pages 99–108. ACM, 2010.

[23] J. Zhao, P. Forer, and A. S. Harvey. Activities, ringmaps and geovisualization of large human movement fields. *Information Visualization*, 7(3-4):198–209, 2008.